AD-758 050

PRINCIPLES OF AUTOMATING CONTROL SYSTEMS

A. N. Romanov, et al

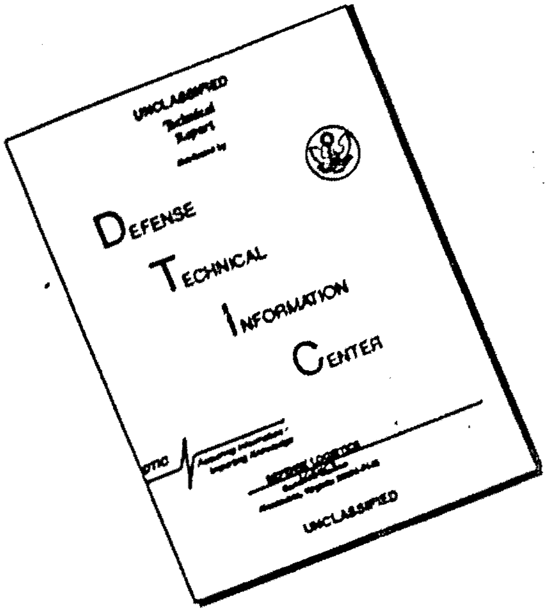
Foreign Technology Division
Wright-Patterson Air Force Base, Ohio

31 January 1973

DISTRIBUTED BY:

NTIS
National Technical Information Service
U. S. DEPARTMENT OF COMMERCE
5285 Port Royal Road, Springfield Va. 22151

DISCLAIMER NOTICE



UNCLASSIFIED
Technical
Report
distributed by

DEFENSE
TECHNICAL
INFORMATION
CENTER

DTIC

UNCLASSIFIED

THIS DOCUMENT IS BEST QUALITY AVAILABLE. THE COPY FURNISHED TO DTIC CONTAINED A SIGNIFICANT NUMBER OF PAGES WHICH DO NOT REPRODUCE LEGIBLY.

AD 758050

FTD-MT-24-1341-72

FOREIGN TECHNOLOGY DIVISION

PRINCIPLES OF AUTOMATING CONTROL SYSTEMS

by

A. N. Romanov and G. A. Frolov

DDC
APR 11 1973

Reproduced by
NATIONAL TECHNICAL
INFORMATION SERVICE
U S Department of Commerce
Springfield VA 22151

Approved for public release;
distribution unlimited.

363

EDITED MACHINE TRANSLATION

FTD-MT-24-1341-72

PRINCIPLES OF AUTOMATING CONTROL SYSTEMS

by: A. N. Romanov and G. A. Frolov

English pages: 351

Source: Avtomatizatsii Sistemy Upravleniya,
Izd vo Voyennoye, Moscow, 1971,
pp. 1-248

Requester: FTD/PDJS

This document is a SYSTRAN machine aided translation, post-edited for technical accuracy by:
SSgt René E. Courville and Robert D. Hill

Approved for public release;
distribution unlimited.

THIS TRANSLATION IS A RENDITION OF THE ORIGINAL FOREIGN TEXT WITHOUT ANY ANALYTICAL OR EDITORIAL COMMENT. STATEMENTS OR THEORIES ADVOCATED OR IMPLIED ARE THOSE OF THE SOURCE AND DO NOT NECESSARILY REFLECT THE POSITION OR OPINION OF THE FOREIGN TECHNOLOGY DIVISION.	PREPARED BY: TRANSLATION DIVISION FOREIGN TECHNOLOGY DIVISION WP-AFB, OHIO.

I

DOCUMENT CONTROL DATA - R & D

(Security classification of title, body of abstract and indexing annotation must be entered when the overall report is classified)

1. ORIGINATING ACTIVITY *(Corporate author)*
Foreign Technology Division
Air Force Systems Command
U. S. Air Force

2a. REPORT SECURITY CLASSIFICATION
UNCLASSIFIED

2b. GROUP

3. REPORT TITLE
PRINCIPLES OF AUTOMATING CONTROL SYSTEMS

4. DESCRIPTIVE NOTES *(Type of report and inclusive dates)*
Translation

5. AUTHOR(S) *(First name, middle initial, last name)*
A. N. Romanov, G. A. Frolov

6. REPORT DATE
1971

7a. TOTAL NO. OF PAGES
351

7b. NO. OF REFS
28

8a. CONTRACT OR GRANT NO.

b. PROJECT NO. AP9G

c.

d. T71-05-10

9a. ORIGINATOR'S REPORT NUMBER(S)
FTD-MT-24-1341-72

9b. OTHER REPORT NO(S) *(Any other numbers that may be assigned this report)*

10. DISTRIBUTION STATEMENT
Approved for public release; distribution unlimited.

11. SUPPLEMENTARY NOTES

12. SPONSORING MILITARY ACTIVITY
Foreign Technology Division
Wright-Patterson AFB, Ohio

13. ABSTRACT
15

DD FORM 1 NOV 65 1473

Table of Contents

U. S. Board on Geographic Names Transliteration System.... vi

Designations of the Trigonometric Functions.............. vii

Preface.. ix

CHAPTER 1. PRINCIPLES OF AUTOMATING THE CONTROL OF AIR DEFENSE FORCES.................................. 1

1.1. General Concepts of Control................. 1

1.2. Essence of Controlling the Combat Operations of Troops............................ 4

1.3. Requirements Set Forth for Systems Controlling Combat Operations of Air Defense Forces.. 8

1.4. Characteristics of Air Defense Forces and the Need for Automating the Control of Their Combat Operations.................................. 13

1.5. Basic Principles of Forming Automated Control Systems.................................. 18

1.6. Classification of Algorithms Which Endure the Use of Computers to Control the Combat Operations of Forces............................ 25

CHAPTER 2. BASIC DATA FROM THE STATISTICAL DECISION THEORY IN CONNECTION WITH THE DETECTION OF SIGNALS.. 28

2.1. Reception of Signals as a Statistical Problem... 28

2.2. Density of Probability Distribution of the Amplitude and Phase of Noise..................... 33

2.3. Mathematical Description of the Interaction of a High-Frequency Signal and Noise.............. 41

2.4. Probability Characteristics of the Quality of Radar Detection................................ 47

2.5. Selection of the Criteria for the Quality of $P_{по}$ and $P_{лт}$ Detection.......................... 52

CHAPTER 3. INITIAL PROCESSING OF RADAR DATA.............. 60

3.1. Determination and Complement of the Initial Processing of Radar Data.......................... 60

3.2. Structural Diagram for the Optimal Detection of a Single Pulse Signal................ 62

3.3. Resolver for Separating a Packet of Pulse Signals.. 65

3.4. Digitization of the Surveillance-Pulse Radar Signals....................................... 70

3.5. Binary Quantization of Signals.............. 75

3.6. Algorithm for the Optimal Detection of a Quantized Signal Packet............................ 79

3.7. Principle Weightless Processing of Radar Signals.. 84

3.8. Methods of Estimating Target Azimuth During Weightless Processing............................ 88

CHAPTER 4. AUTOMATIC EXTRACTION AND CODING OF TARGET COORDINATES.................................... 90

4.1. Conjugation of Radar with Computers......... 90

4.2. Digital Device for Detecting Radar Signals (Preselector).................................... 93

4.3. Ranging in Digital Code..................... 105

4.4. Digital Measurement of Azimuth.............. 111

4.5. Algorithm for the Primary Processing of Radar Information................................ 119

4.6. Equipment for the Semiautomatic Extraction of Information.................................. 127

CHAPTER 5. SECONDARY PROCESSING OF RADAR INFORMATION..... 131

5.1. Definition and Makeup of Secondary Processing.. 131

5.2. Construction of the Trajectory of Target Movement.. 135

5.3. Automatic Detection (Automatic Lock-On) of Trajectories................................ 139

5.4. Extrapolation of a Target Trajectory and its Smoothing.. 143

5.5. Tracking the Trajectories of Targets........ 153

5.6. Gating and Selection of Blips in Gates...... 155

5.7. Criteria for the Quality of Secondary Processing.. 164

5.8. Algorithm for the Secondary Processing of Radar Information.................................. 169

CHAPTER 6. TERTIARY DATA PROCESSING.................... 173

6.1. Definition and Makeup of Tertiary Processing 173

6.2. Assembling Reports Incoming from Radar Stations.. 174

6.3. Reduction of Target Locationg Blips to a Single System of Coordinates..................... 180

6.4. Reduction of the Target Locationg Coordinates to a Single Reference Time........... 186

6.5. Identification of the Blips of Target Location.. 188

6.6. Averaging Target Location Blips............ 195

CHAPTER 7. BASIC INFORMATION OF THE THEORY OF MESSAGE TRANSMISSION IN A COMMUNICATION CHANNEL....... 198

7.1. Basic Problems of the Theory of Message Transmission.. 198

7.2. Concept of Information..................... 203

7.3. Representation of Messages and Evaluation of the Quantity of Information in Equiprobable States of Elements of Messages 205

7.4. Determination of the Quantity of Information in the Unequally Probable States of Elements of the Messages 212

7.5 Coding of the Messages 218

7.6. Conditional Entropy of Statistically Dependent Messages 224

7.7. The Transmission of Discrete Messages Along the Communication Channels with Noises and Without Noises 227

7.8. Formation of Binary Signals in the Automated Control Systems 235

7.9. Features of the Construction of Discrete Communication Systems Which use Uniform Codes 240

7.10. Layout and Operating Principles of One of the Existing Systems of Data Transmission to the Automatic Control System 246

CHAPTER 8. METHODS AND MEANS OF THE DISPLAY OF INFORMATION 254

8.1. Methods of Visual Display of Information 254

8.2. Methods of the Formation of Elements of the Image in the Electronic Devices of Displaying Graphical Information 257

8.3. Electromechanical Devices of the Output of Graphical Information 275

8.4. Electroluminescent Devices for the Output of Graphical Information 278

8.5. Photochemical and the Electrochemical Devices for the Output of Graphical Information ... 281

8.6. Basic Requirements for the System of the Graphic Display of Information 289

CHAPTER 9. DEVELOPMENT OF THE SOLUTION FOR COMBAT OPERATIONS OF ACTIVE MEANS 293

9.2. Basic Principles of Target Assignment....... 296

9.3. Optimization of the Index of Effectiveness by Means of Methods of Linear Programming......... 305

9.4. Solution of Target-Assignment Problems by the Hungarian Method............................. 318

CHAPTER 10. COMBAT EMPLOYMENT OF AUTOMATIC CONTROL SYSTEMS IN AIR DEFENSE FORCES................ 327

10.1. Basic Functions Being Fulfilled by the Combat Crews of Control Posts Equipped with ACS.... 327

10.2. Removal and Initial Processing of Information on the Radar Connected with the ACS.... 330

10.3. Information Processing and the Control Combat Means of Control Posts of the ACS........... 333

10.4. Brief Survey of Some ACS of Capitalist Countries...................................... 343

Bibliography.. 351

U. S. BOARD ON GEOGRAPHIC NAMES TRANSLITERATION SYSTEM

Block	Italic	Transliteration	Block	Italic	Transliteration
А а	*А а*	A, a	Р р	*Р р*	R, r
Б б	*Б б*	B, b	С с	*С с*	S, s
В в	*В в*	V, v	Т т	*Т т*	T, t
Г г	*Г г*	G, g	У у	*У у*	U, u
Д д	*Д д*	D, d	Ф ф	*Ф ф*	F, f
Е е	*Е е*	Ye, ye; E, e*	Х х	*Х х*	Kh, kh
Ж ж	*Ж ж*	Zh, zh	Ц ц	*Ц ц*	Ts, ts
З з	*З з*	Z, z	Ч ч	*Ч ч*	Ch, ch
И и	*И и*	I, i	Ш ш	*Ш ш*	Sh, sh
Й й	*Й й*	Y, y	Щ щ	*Щ щ*	Shch, shch
К к	*К к*	K, k	Ъ ъ	*Ъ ъ*	"
Л л	*Л л*	L, l	Ы ы	*Ы ы*	Y, y
М м	*М м*	M, m	Ь ь	*Ь ь*	'
Н н	*Н н*	N, n	Э э	*Э э*	E, e
О о	*О о*	O, o	Ю ю	*Ю ю*	Yu, yu
П п	*П п*	P, p	Я я	*Я я*	Ya, ya

* ye initially, after vowels, and after ъ, ь; e elsewhere. When written as ё in Russian, transliterate as yё or ё. The use of diacritical marks is preferred, but such marks may be omitted when expediency dictates.

FOLLOWING ARE THE CORRESPONDING RUSSIAN AND ENGLISH DESIGNATIONS OF THE TRIGONOMETRIC FUNCTIONS

Russian	English
sin	sin
cos	cos
tg	tan
ctg	cot
sec	sec
cosec	csc
sh	sinh
ch	cosh
th	tanh
cth	coth
sch	sech
csch	csch
arc sin	$\sin^{-1}$
arc cos	$\cos^{-1}$
arc tg	$\tan^{-1}$
arc ctg	$\cot^{-1}$
arc sec	$\sec^{-1}$
arc cosec	$\csc^{-1}$
arc sh	$\sinh^{-1}$
arc ch	$\cosh^{-1}$
arc th	$\tanh^{-1}$
arc cth	$\coth^{-1}$
arc sch	sech^{-1}
arc csch	csch^{-1}

---

rot	curl
lg	log

Principles of Automating Control Systems

M., Voyenizdat, 1971

This book is one of the first works in which the principles of forming automated systems for controlling the combat operations of air defense forces are systematically presented using modern mathematical methods. It is written using the materials of the open domestic and foreign press.

The book examines the problems of algorithmizing the processes of radar data processing and the methods of realizing these algorithms with the aid of digital computers. Considerable attention is paid to methods of mapping and transmitting information over communication channels, as well as to problems in the practical performance of the basic elements of automated control systems [ACS] (АСУ). The stated material is explained by examples.

The book has been dependent on engineers who deal with the development, testing and operation of ACS in air defense forces. It can be used by the cadets of military academies, by academy students and by the students of technical colleges for study on the principles of ACS formation.

PREFACE

In recent years requirements for the operational efficiency of controlling the combat operations of the air defense forces have been continuously rising in connection with the increase in weapon power and in speeds of delivering them to attack objectives. These requirements are becoming all the more urgent, that the shortage of time to use active combat means to intercept and destroy aerial targets is felt all the more distinctly.

The accomplishment of the indicated requirements became possible because of wide adoption by the air defense forces of digital computer technology which in turn led to the creation of numerous automated control systems [ACS] (АСУ). The operation of ACS requires of specialists a knowledge not only of the principles of radar, electronics and computer technology, but also a knowledge of the queueing theory, the theory of games, linear and dynamic programming and the statistical decision theory.

This book attemps to systematize the materials on the principles of ACS formation.

Chapters 1-8 and 10 are written by A. N. Romanov and G. A. Frolov. Chapter 9 is written by M. M. Shokurov.

CHAPTER 1

PRINCIPLES OF AUTOMATING THE CONTROL OF AIR DEFENSE FORCES

1.1. General Concepts of Control

In general, control is that organization of one or another process which ensures the achievement of determined objectives. Development in electronics and computer technology led to the incorporation of automation into almost all areas of human activity. Today, control engineering automation are the principal direction in the development of all technology. The era of creating machines which perform certain functions of the human brain, in particular those which control various complex processes, has come.

For a better understanding of the basic principles of control let us examine, for example, the process of controlling an aircraft. The pilot sees an instrument panel in front of him and determines course, speed, and flight altitude by the instruments. He decides to change any of the parameters on the basis of these data and the assigned mission. By analyzing the control process, it is possible to distinguish the following basic elements.

- **information on the problem of control** (in our example, the information on the required heading, speed and flight altitude

of the aircraft) and **on the results of the control** (the pilot sees, observing the instruments, where, at what altitude and with what speed the aircraft he is piloting is flying);

- **analysis of the information** obtained by him and on the basis of this making a decision on the needed control actions;

- **performing** the made decision.

The three elements indicated comprise the principle of any control. In accordance with this, to organize the control process it is necessary to have sources of information on the problems of control and the results of control, units for analyzing the obtained information and developing a solution, and also actuating units for controlling an object.

In the organization of the control process a large (and sometimes a decisive one too) role is played by the information on the results of the control. With this information the decision on the controlling actions depends substantially on the results of the control. A closed circuit results: the reason which causes the change in status of the control object is placed in dependence on what kind of result it causes. Such a connection of cause and effect is called feedback. The principle of control using feedback underlies the overwhelming majority of control processes.

In order that the information can be used in the control process, in the majority of cases it must acquire a waveform. The signal form of the information is distinctive characteristic of the control processes. In other words, it can be said that control is the conversion of the information into signals which correct the activity of the control object.

The combination of all the units which ensure the control of any object is called the control system. If the functions of all the elements of a control system are performed by various systems without the direct participation of a man, then system is called automatic. For example, the autopilot which controls the flight of an aircraft, the system for guiding a surface-to-air guided missile to a target, the automatic tracking of an aerial target by a radar station, etc., can be attributed to such systems.

In a number of cases the complex control process is performed by a man (sometimes a group of people) with the aid of various automatic systems. Such a control system is called automated. A system for controlling the combat operations of troops, which contains a complex of automatic systems, computers and personnel, which gather and process information and which develop the initial data for decision making by the commander, can serve as an example of an automated control system [ACS] (АСУ).

When studying the control process it is necessary to examine joint operation, as well as the interaction of the control system and the control objects, which form a complex dynamic system.

In some cases the problem of control is the guarantee of stability of some physical quantity. This particular form of control is called regulation. An automatic control system which ensures regulation is called a regulator, and a system consisting of a regulated object and a regulator is called an automatic regulation system.

Summing up what has been said, it can be said that any control is accomplished on the basis of regulating the process of converting the information on the status and functioning conditions of objectives into signals which guarantee either retaining the status of the object or reducing it in conformity with the program.

The control processes are inherent to systems of biological and social order, as well as to systems artificially created by man (machines, mechanisms). It is acceptable to distinguish three basic areas, spheres of control: controlling the activity of human associations which decide one or another problem; controlling the processes which occur in living organisms; controlling systems of machines, technological processes and processes in general which take place under the purposeful influence of man on nature.

The control process is considered to be studied, if it is possible to reproduce all the conversions which occur in a system under the effect of control signals, as well as to reproduce the sequence of developing control signals. Such a system of instructions, which is expressed formally, is called the algorithmic description of the control processes. The algorithmization of the control processes is one of the major problems, since it makes it possible to develop optimum methods of control.

.2. Essence of Controlling the Combat Operations of Troops

Continual supervision directed toward the accomplishment of assigned missions on the part of the commanders and echelons of the activity of troops subordinate to them is what is meant by controlling the combat operations of troops.

Controlling troops in combat consists of the timely and steady execution of measures which ensure thorough training, the careful organization and the successful performance of combat missions by all subordinate forces. Troop control includes the following stages: gathering information on the situation, evaluating it and drawing conclusions in the light of the assigned mission, making calculations, making decisions and bringing them to the executors, checking the actions of the executors, and adjusting decisions when the situation changes. Consequently, all the basic processes of controlling troops are connected with gathering, processing and transmitting information.

The information necessary for troop control is called operational-tactical information, which is usually divided into three groups.

The first group is comprised of data on the enemy, the status of their forces, etc. This information, which characterizes the specified current status of the system is status information.

The second group includes the rules by which the status information is processed. In practice, this is the tactics stated in manuals and regulations, and the combat experience of the commander making the decision.

Finally, the third information group is comprised of the decisions which the commander makes on the basis of analysis of the status information. These decisions or commands are called control actions or command information.

Thus, troop control during combat operations can be defined as the conversion of status information into command information conformity with the algorithms of control.

The closed control loop in the military system is given in Fig. 1.1. The control element here is the commander and his staff; the control objects - the subordinate forces; the environment - the enemy and other factors of the situation (for example, terrain, meteorological conditions, etc.).

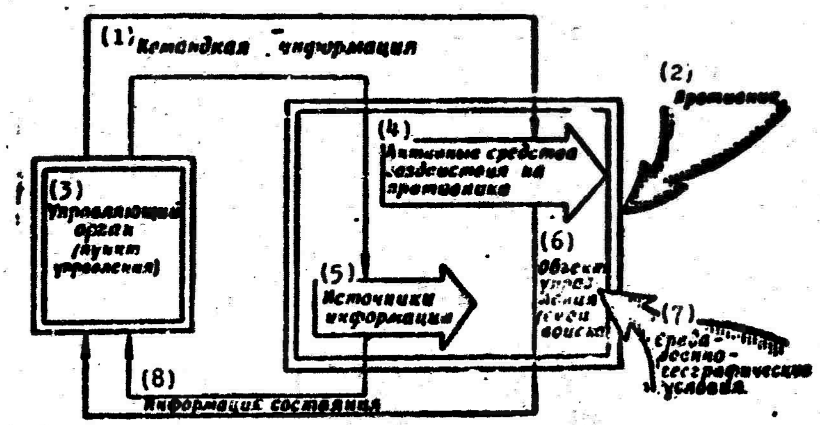

(1) Командная информация
(2) Противник
(3) Управляющий орган (пункт управления)
(4) Активные средства воздействия на противника
(5) Источники информации
(6) Объект управления (свои войска)
(7) Среда — военно-географические условия
(8) Информация состояния

Fig. 1.1. Closed control loop in the military system.
KEY: (1) Command information; (2) Enemy; (3) Control element (control post); (4) Active means of action against the enemy; (5) Sources of information; (6) Control object (own forces); (7) Environment - military geographical conditions; (8) Status information.

By a complex dynamic system is meant a closed-loop system made up of interconnected and interacting control elements and control object. A control system can be multistage and consist of a number of subsystems.

If we examine an air defense system as dynamic (Fig. 1.2), its subsystem will be the air defense areas which in turn are divided into air defense sectors with operational guidance centers. The active means of air defense (radar aids - radar sites, interceptors, and SAM sites) will be the control objects.

The operation of this system can in the first approximation be described in the following manner. Radar aids acquire information on the air enemy and transmit it to their operational guidance center, from which the information in sequence goes into the air defense area and finally to the control post [CP] (ПУ) of the operational center of NORAD (North American Air Defense Command).

At the control post the information obtained from all radar aids is generalized and a decision is made as to what active means in a particular air defense area should destroy the aerial targets. The appropriate command information is transmitted to the selected air defense area.

The decision (what kind of or by which air defense sectors will the assigned aerial targets be destroyed) is made in the operations center of the air defense area and the appropriate commands are issued at the operational guidance centers of these sectors.

The specific active means which will be directed to the aerial target (interceptor guidance or acquisition SAM guidance radars) is assigned at the operational guidance centers.

Figure 1.2 graphically shows the multistage structure with the rather complex connections.

The sharp state of conflict of situations and the armed conflict of two enemies, which is controlled with determined goals, are the basic feature of military dynamic systems, which distinguish them from the dynamic systems of other areas of application. During the control of combat troop operations special importance is gained by the feedback which makes it possible to adjust control actions, depending on the nature of actions of the enemy.

Hence, the essence of controlling combat troop operations can be formulated as a change in the input parameters of the military dynamic system to purposefully change its status.

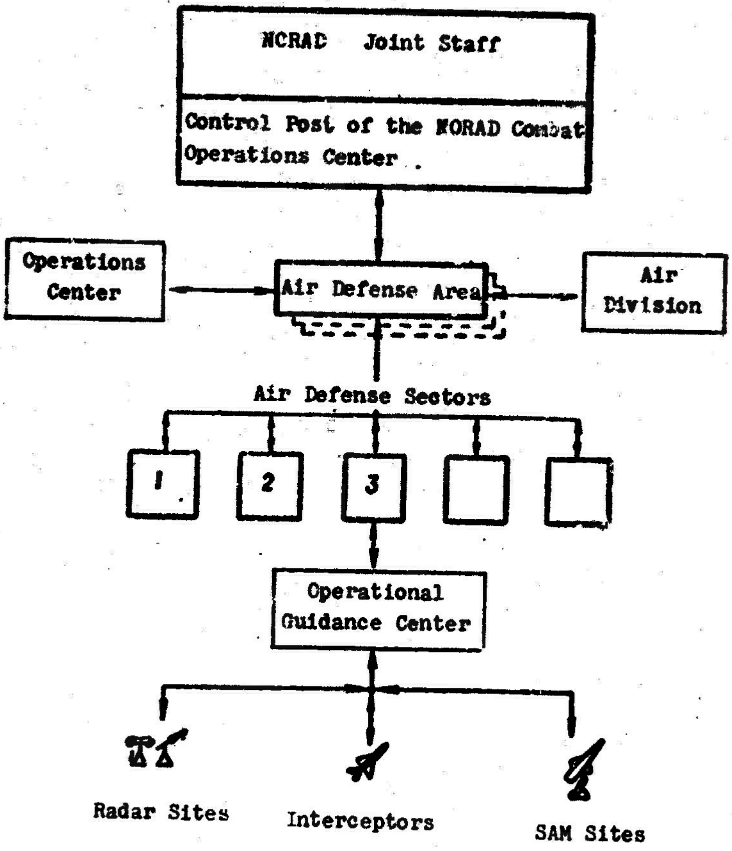


Fig. 1.2. The structure of the air defense system of the USA.

1.3. Requirements Set Forth for Systems Controlling Combat Operations of Air Defense Forces

As is known, the basis of the striking powers of the armies of imperialistic nations is strategic missiles. At the same time the improvement of piloted air attack weapons continues abroad and new methods of using nuclear missile weapons are practically mastered.

The future development of air attack weapons (increasing the speeds, ceiling and flight range of strategic bombers, the capability for mass air attack, etc.) and the tatics of their use require the development of new methods and means of operational troop control, which correspond to modern methods of conducting combat operations.

A study of the materials on military training exercises being carried out by the armies of imperialistic nations, the reports of military specialists, and analysis of their tendencies in modern armament development reveal the features of future combat operations, which can be determined by the suddenness, the decisive actions and the high fluidity of operations.

The basic requirement set forth for combat operations control systems is the timeliness of decision making for directing the active means and transmitting commands to the executors. The time spent on the control and execution of commands can be presented in the form of summing up the following components:

$$T_{\text{упр}} = t_{\text{обр}} + t_{\text{реш}} + t_{\text{пер}} + t_{\text{гот}}$$

where $T_{\text{упр}}$ - the time spent on the control and execution of commands;

$t_{\text{обр}}$ - the time of processing and transmitting status information (for example, processing the signals of radar sites);

$t_{\text{реш}}$ - the time necessary to understand the situation and to develop a solution;

$t_{\text{пер}}$ - the time of forming and transmitting command information to the executors;

$t_{\text{гот}}$ - the readiness time for active means and the time they take to perform the necessary actions.

The first three terms comprise the action time of the control system, and the fourth - the action time of the executive agencies. With constant $T_{\text{упр}}$ an increase in the action time of the control system reduces preparation time and action time of the active means, which in a number of cases is inadmissible.

Hence, the need follows to increase the high-speed control systems. This is reached, in particular, by automating it.

The systems for controlling troops should ensure timeliness in the inflow of information about environmental conditions and events, since the problems of gathering and processing the information and transmitting it to the different links in the military control systems are paramount for the control process.

Control systems must also ensure rapid analysis of the situation and making the necessary calculations for decision making. Furthermore, control systems must rapidly bring the decision made by the commander to the executors and must continuously monitor the actions of the controlled forces and means.

Hence, troop control systems have the following basic requirements:

- the capability to timely bring all elements of the system into combat readiness, beginning with lowest and ending with the highest control link, in order to ensure the execution of the troop combat missions;

- solving control problems with the sufficiently effective realization of them by troops in the shortest possible time;

- the high reliability of the control system, its ability to ensure continuous control of forces under any conditions;

- the quickness in reacting to a change in the situation;

- the capability of ensuring the efficiency of the control system upon the creation by the enemy of a different kind of jamming, i.e., the ability of the system to separate the useful information against the background of strong interferences in an amount sufficient to decide combat missions;

- the close interconnection between the different elements of the system and the capability of interaction between them; this to a considerable degree determines the viability of the control system and makes it more flexible when solving the problems of target-distribution and destruction of aerial targets.

Let us analyze these requirements in the example of an air defense system. Upon an air enemy attack the following must be brought into combat readiness: all links of the system, beginning with the lowest and ending with the highest, i.e., the entire radar system from the individual radar stations to the largest points for gathering and processing radar information; all active facilities, beginning from the individual antiaircraft-missile and fighter units and ending with the largest control posts which coordinate conducting combat operations in the entire territory being defended.

All of these links must begin to function before the attacking side can begin to perform its combat mission (i.e., before, for example, enemy aircraft can launch missiles of the "air-to-ground" class, or before bombs can be dropped from aircraft).

In this case it is necessary to bear in mind that the chief purpose of air defense forces is the destruction of an air enemy even before it performs its combat mission. Thus, for timely destruction of an air enemy it is necessary to first of all organize air reconnaissance to a specified depth. Thus, the minimum reconnaissance depth for antiaircraft missile units must ensure the detection of aerial targets at distances which make it possible to determine the national identity of the targets and their characteristics, to evaluate the air situation, to make a timely decision for conducting combat operations and to state the mission to subordinate units for the destruction of the targets at the far boundary of the strike zone.

A successful solution to the problems of control primarily depends on whether the control system is automated or not and what the degree of its automation is. Undoubtedly, with an automated system the time for solving the problems of control is reduced considerably.

The requirement for the high invulnerability of the system to jamming requires such information processing during which the ability of the system to extract the useful information against the background of strong interferences would be sufficient for deciding the combat missions.

The need for the cooperation of the different elements of the system is easy to illustrate by a simple example.

The defended object is attacked by two bombers. The object is defended by antiaircraft missiles and interceptors. In this case the missiles and fighters can simultaneously destroy only one bomber. The mission is that of rationally dispersing the forces of the antiaircraft-missile and fighter unit to the two targets, i.e., concretely determining which combat device, when, on which line of defense should a specific target be destroyed while not alloying the bomber to get to the objective.

The performance of such a task is impossible without the close cooperation of subdivisions or units of antiaircraft missiles and fighters, which should be carried out under the direction and control of a superior commander. Without the effective control of the different forms of combat devices a modern troop control system is unthinkable, since the lack of effective control, as a rule, leads to a disruption in performing the combat mission.

1.4. Characteristics of Air Defense Forces and the Need for Automating the Control of Their Combat Operations

Air defense forces are called to achieve reliable defense of the airspace of an area of the country. The forces perform their own missions, destroying the aircraft and pilotless air attack weapons of the enemy.

An intricate complex of various forces and means with different peculiarities and capabilities is required to successfully combat an air enemy. This is necessary, because the process of warding off an air attack includes a number of actions (different in nature and methods of execution), such as detecting an air enemy, identifying it, determining the current coordinates of the target and the parameters of their movement, lock-on and tracking of the targets by radar sites and finally destroying them. In accordance with this, the air defense forces consist of units and subdivisions of antiaircraft missile, interceptors, radio aids and means of communication.

Antiaircraft missiles are intended for the destruction of an air enemy on routes of approach to the defended objects. They are the basis of air defense. Antiaircraft-missile units and subdivisions are armed with antiaircraft missile complexes, capable with sufficiently great effectiveness to destroy modern piloted and pilotless air attack weapons both on near and on remote routes of approach to the defended objects over a wide range of altitudes.

Interceptors are also intended for the destruction of an air enemy on routes of approach to the covered areas and objectives. They are armed with "air-to-air" missiles. Interceptors are the most maneuverable means of the air defense forces.

Radio aids are intended for the detection of air targets, their identification, determining their coordinates and their parameters of movement, as well as for the radar support of antiaircraft-missile and fighter units. To perform their tasks they have radar detection aids, as well as means of warning, controlling and communication.

Success in warding off air attack weapons depends much on timely detecting them, warning the air defense forces about the beginning of an attack, rapidly getting them into readiness to ward off the air penetration and the continuous control of the combat operations of units and subdivisions.

The latter is accomplished only with the maximally possible automation of the processes of gathering and processing information at command posts, as well as automation of the processes of controlling the active facilities. In this case by control are meant all processes connected with obtaining data for developing and making a decision by the commander and the measures taken to realize the made decision.

The need for automating the majority of operations in gathering and processing information, as well as for automating the processes of controlling the combat operations of the active means of air defense, is determined by the following factors.

Air defense forces combat the most mobile offensive means having high flight speeds. Therefore, in order to be able to defeat an enemy even on the routes of approach to the covered areas or objectives, the active means of the air defense forces must derive the combat missions in proper time.

Air attack weapons have a long range. This makes it possible for them to concentrate forces in one or several directions.

When there are a large number of enemy aircraft and our own interceptors in the air and the combat operations are short-lived, only the use of automated control systems can provide the commander with sufficiently reliable data on the air situation, data which are necessary to him for timely adoption of expedient decisions and to rapidly bring the combat missions to the forces.

There are means of automating control in all branches of the air defense forces. Even prior to the Second World War antiaircraft artillery used Antiaircraft Fire Directors [AAFD] (ПУАЗО). In antiaircraft missile units large part of the operations connected with combat control is automated, and the missile guidance process is completely automated.

The means of automating control penetrated comparatively slowly into the radio engineering forces. This is explained by the complexity of problems in the automatic extraction, transmission and processing of radar information.

Requirements for reducing the time of obtaining the initial data for decision making and for their authenticity have increased. However, the capabilities of man are limited. In the absence of control automation means the functioning of posts for controlling troops was based mainly on the mental activity of combat crews. As is known, this mental activity is not rapid enough. An increase in the operating speeds of control post crews and in their handling capacity, true only up to a certain limit, is possible on account of the long training sessions and the increased numerical strength of combat crews. The radical improvement in the operation of control posts became possible with incorporation into the forces of electronic computers. In this case the most important elements of the combat operation are formalized (they are mathematically expressed in the form of algorithms); the programs are composed according to the algorithms. Computers run the programs, giving upon output the solution for approval by the commander.

In connection with this, there is a need for various units which ensure the obtaining of initial data and their input into the computer, as well as units for the output and mapping of results in a form convenient for rapid perception by the commander.

The need for the wide complex automation of controlling the combat operations of the air defense system becomes clearer, if we enumerate the problems being solved in the process of mapping an air attack.

This situation can be simply presented in the example of covering a single objective by an antiaircraft missile unit and by interceptors. To ensure the disposition of air defense means, the Americans usually make the following calculation.

Radar detects aerial targets at a distance of 500-600 km. In this case the detection and recognition expends up to 7 min. In this same time the signal can be conveyed to fighter aviation airfields. At the 400 km line the antiaircraft missile units are warned. At the 250-300 km line the targets can be attacked by interceptors. Antiaircraft missiles are launched with such calculation that the rendezvous of the missile with the target will take place at the distant boundary of the kill zone of the antiaircraft missile site.

It is clear that the time available at the disposal of the commander, especially if the target is fast, is extremely limited. Without determining the measures which sharply reduce wasting the time to perform assigned missions, the overall mission - the destruction of aircraft at established lines - can be performed.

As one of such measures it is advantageous to use a high-speed automatic (or automated) control system.

In summary, the following conclusions can be drawn.

1. An air defense system is an intricate complex structure in which the numerous information, material and energetic flows requiring coordination and regulation with a speed and accuracy sometimes inaccessible for man are interwoven.

2. The automation of control should reduce the time for transmitting and processing the information, the time for evaluating the air situation and making a decision; it should automate target identification and guidance of active means, as well as to reduce the periods for bringing the made decisions to troops and of obtaining information about their combat operations.

Hence, it is primarily advantageous to automate the following processes:

- the extraction and transmission of data on the air situation from radar stations;

- the reception and processing of radar information from various sources;

- the derivation of information to graphically represent the air situation and the dynamics of combat;

- the gathering, generalization and mapping of information on the combat readiness of forces;

- the development of initial data to make a decision on warding off the air penetration of an air enemy;

- target identification and the guidance of active means.

1.5. Basic Principles of Forming Automated Control Systems

Each of the just enumerated processes can be automated according to their own peculiarities with the aid of various technical equipment. In this case a definite effect in refining the quality of control can be achieved.

It is natural that the greatest effect from the automation of control will be obtained when the automation of all control processes will be interconnected in such a way that one process proceeds into another, without creating difficulties and inhibitions. Only in this case will the automation of control bring the greatest success.

The control automation with which all the basic control processes are automated to a necessary degree and their automation is mutually connected is called complex automation. Precisely, such automation of the processes is progressive, facilitating increased control effectiveness.

It is considered to be necessary to develop complex automated control systems to achieve complex automation for controlling military dynamic systems.

By a complex automated control system is meant the set of technical equipment and personnel united in a complex of the interrelated control loops and providing the necessary tempos and the spatial coverage of the sphere of control on the basis of the coordinate automation of the control processes according to the accepted organization of its forces and their control scheme.

It has the following elements in its composition:

- means for the extraction of information;

- communication channels which ensure the transmission of information from its extraction sources to processing points;

- facilities of control posts which generate the processing of information and the development of control actions - commands;

- communication channels which ensure the transmission of command information to the objects of control and information on their status to the control posts;

- control posts for the subordinate forces.

The simplified circuit for controlling the "Sage" air defense forces of the USA can be given as an example of a complex automated control system.

The composition of the system (Fig. 1.3) includes numerous and various devices pertinent to the control technique: radar sets, means of communication, electronic computers, and equipment for conversion, coupling and data input from the means of extracting the information into electronic computers, graphic display equipment (screens, plotting boards and indicators), different commutating equipment and finally power sources.

High-speed computers process status information into command information. They are the "brain" of all automated control systems, and only with their appearance did it become possible to create similar systems. However, it must be emphasized that some high-speed computers are insufficient for the creation of complex automated control systems. Without the various means of obtaining information (to them refer various range-only radars, altimeters, visual reconnaissance posts and others),

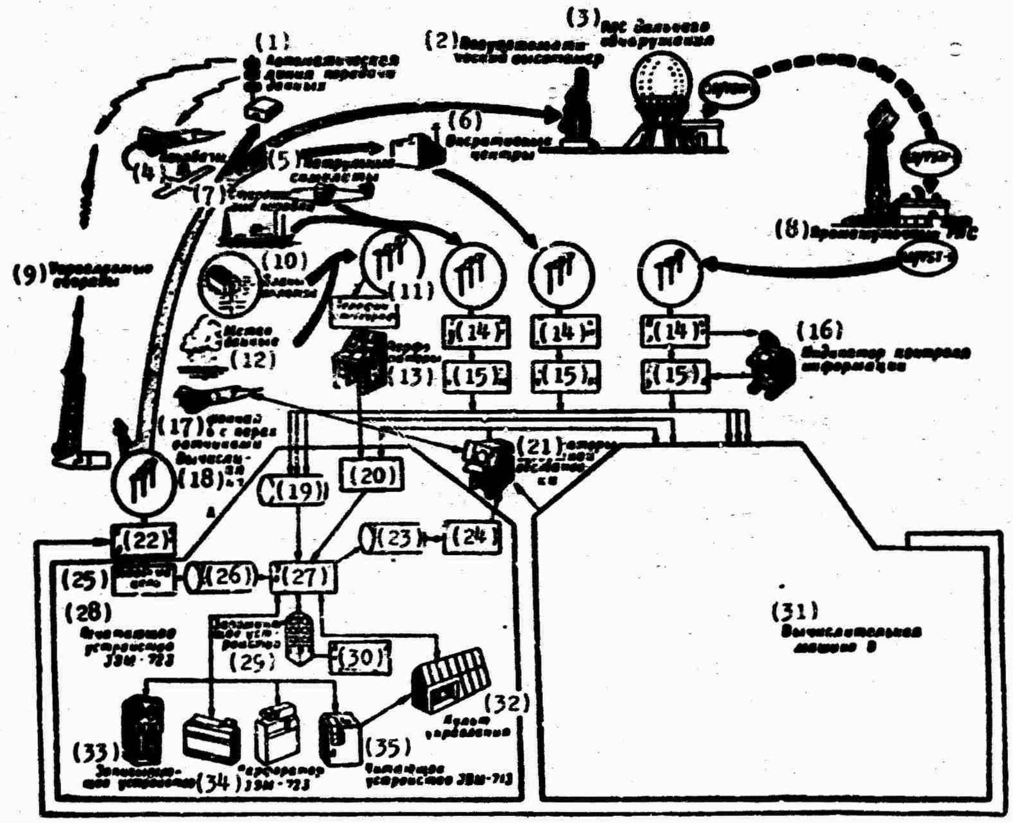


Fig. 1.3. The simplified control system of the air defense forces "Sage" of the USA.
KEY: (1) Automatic data transmission link; (2) Semi-automatic altimeter; (3) Long-range detection radar stations; (4) Air bases; (5) Patron aircraft; (6) Operational centers; (7) Patrol vessels; (8) Intermediate radar stations; (9) Guided missiles; (10) Flight plans; (11) Telephone and telegraph; (12) Meteorological data; (13) Keypunch machines; (14) Numerical data receiver; (15) Input equipment; (16) Information monitoring display; (17) Telephone communication with interceptors; (18) Computer; (19) Input drums; (20) Manual input; (21) Air situation displays; (22) Numerical data transmitter; (23) Mapping drums; (24) Mapping generators; (25) Output circuits; (26) Output drums; (27) Control equipment; (28) EVM-723 printer; (29) Memory unit; (30) Arithmetic unit; (31) Computer V; (32) Control panel; (33) Recorder mechanism; (34) EVM-723 Keypunch machine; (35) EVM-713 character reader.

equipment for coupling them with the communication channels, communication channels, equipment for coupling the communication channels with computers, the creation of a complex automated control systems would be just as impossible as without the computers themselves.

Let us examine a structural diagram of hypothetical automated system for controlling the combat operations of air defense forces (Fig 1.4), which forms a closed control loop.

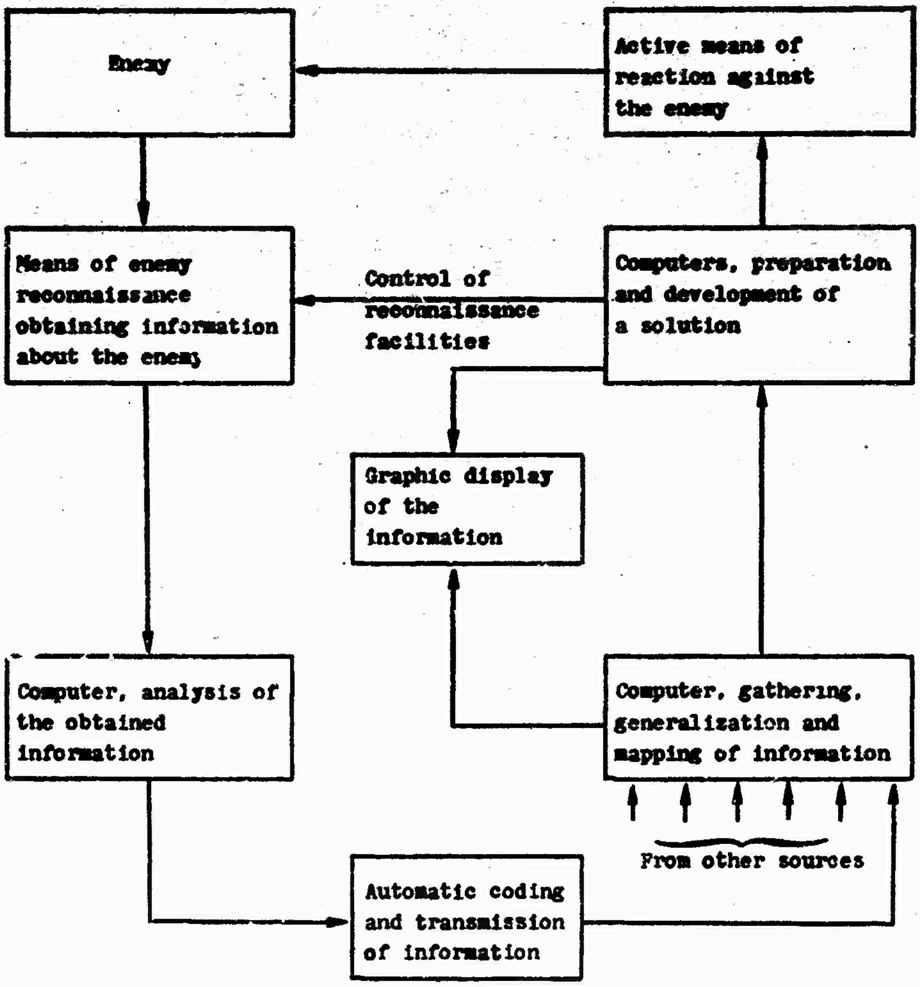

Enemy
Active means of reaction against the enemy
Means of enemy reconnaissance obtaining information about the enemy
Control of reconnaissance facilities
Computers, preparation and development of a solution
Graphic display of the information
Computer, analysis of the obtained information
Computer, gathering, generalization and mapping of information
From other sources
Automatic coding and transmission of information

Fig. 1.4. Structural diagram of the automated system for controlling the combat operations of the air defense forces.

The basic means of extracting information in this system are the radar stations which give out data on enemy targets. In order that the information will be sufficiently reliable and will arrive in proper time, it is necessary to automate extraction of radar information and to convert it a form convenient for computer input. These operations are performed by automatic or semi-automatic equipment in the initial processing of radar information.

The automatic extraction of data from radar sets and their automatic input into a computer fully automates those control circuits into which the radar stations come. Analyzing the information obtained from the radar stations and processing it are accomplished during the so-called reprocessing of radar information. In an ACS a specialized computer, as a rule, does the reprocessing. It performs an extensive cycle of actions connected with the separation of useful signals from interferences and with determination of coordinates, movement parameters and a number of other characteristics of moving targets.

The next stage is coding and transmitting the information. The computer-processed information is coded and proceeds over the communication channel to a specified addressee (to a control post) in the form of a message. As a rule, all the operations performed in this stage are fully automated.

Information from all the units performing reconnaissance of the airspace, and the status information from their combat means is gathered at the control post. All of the information incoming to the control post should be generalized and is graphically represented. The generalization and analysis of the information coming in from various sources includes the tertiary processing of the information.

In the ACS the functions of tertiary processing are performed by a specialized computer located at the control post. This machine must complete the task of: decoding the message; comparing it with messages received earlier; supplementing the information on a given target or discarding this message; bringing the information obtained in the message to a single origin time reading and to a single system of coordinates;

preparing a target data card for delivery to the graphic display unit or for storage in the memory unit.

The computer is capable of distinguishing, separating and sifting out improbable data by specially developed programs. During the composition of such programs, situations where the doubts in the accuracy and authenticity of the incoming and processed information can arise are determined in advance. The rules for isolating similar situations, sifting them out or increasing the authenticity of the information are also included in the program.

Computers can be easily adapted also to display a graphic representation of the situation. They can distribute the information obtained on those or other objectives between the different display units, with consideration of their importance and scale, and they can fashion both symbols for the objectives and explanatory labels for them.

The next stage in the combat operation of evaluating the control post is preparing the development of a decision on controlling active means. The problems of target assignments and target identification are solved here, i.e., active facilities best lock on the targets underlying destruction. The coordinates and parameters of movement for that target which must be destroyed are given to each selected active facility. In this case one or another means of action can be prescribed (the SAM method, the number and type of missiles, the route of interceptor movement to the target, etc.). All of these problems can be successfully solved with the aid of a specialized computer.

The chief advantage of solving problems of this nature on a computer is imparting confidence and validity to the reasonings of the commander during decision-making. If each (or at least in necessary and possible cases) reasoning or assumption of the commander in the course of making a decision will be supported or rejected by the calculations made at a rapid (not delaying the course of the reasonings) rate, this will free the commander from many difficulties and errors.

Another very important advantage of using machines for calculations in the course of preparing a solution is the capability to rapidly perform a large number of alternatives of the same calculation at constant values of the needed parameters. Without such a computer the commander frequently does not have sufficient time to even calculate one (often not the best) alternative. If calculations are conducted by many alternatives, commander can analyze their results and select the best one.

And finally, the last stage of the combat operation of the system is the guidance of active means to an air enemy and controlling them in the course of combat. This stage is directly connected with the development and transmission of command information to air defense forces and facilities.

Here one should first refer to the transmission of selected commands of control over lines of communication. Electronic computers can code the text of the made decision, introduce the command into the link and give a signal upon the transmission of the given command.

Secondly, this stage also includes the direct control of missiles, radars, and interceptors. This control is accomplished with the aid of controlling computers.

1.6. Classification of Algorithms which Endure the Use of Computers to Control the Combat Operations of Forces

The range of control problems which can be solved by computers is constantly being broadened basically as a function of how rapidly it is possible to formalize the process of execution by man of one or another control function, as well as to develop the appropriate algorithm.

The term "algorithm," like the special division of mathematics - the theory of algorithms - existed and was used in mathematics prior to the emergence of cybernetics.

By algorithm in mathematics is meant the precise ordering of the execution in specified order of a certain system of actions leading to the solution of a specified problem.

In automated control systems the algorithm is called the system of formal rules and control conditions whose execution in assigned sequence makes it possible to solve a problem without understanding its essence.

The solution of some intellectual problems is possible on a computer with the aid of algorithms.

After compiling the control algorithms which describe the different methods control, it is possible to analyze these algorithms, and through them, to also select the best control method. Thus, the development of algorithms for controlling the different processes is one of the most important stages in the automation of control systems.

Since human thinking is based not only on formal, but also on dialectical logic, the formalization of the mental activity of

man is very complex. At the present time the operation of automated systems requires strict formalization of the processes occurring in it, i.e., the development of appropriate algorithms. One of the basic difficulties in developing algorithms which simulate human thinking results in this contradiction.

Specific difficulty in developing algorithms for controlling the active means of forces in combat operations is the uncertainty introduced into the control process by the insufficiency of information about enemy actions.

It is presently possible to distinguish the three basic groups of the algorithms which insure the use of computers in controlling the combat operations of air defense forces:

- the algorithms of gathering and processing information;
- the algorithms of preparing and making a decision;
- the algorithms of transmitting information.

The algorithms of gathering and processing information make it possible to use computers to solve problems in the classification of information, in making "data cards" with the characteristics of the targets, in forming images on information display units. This group also includes the algorithms which determine the operation of the computer in sorting, arranging in the memory units and output in proportion to necessity of information which is available in a given control member.

The algorithms of preparing and developing a solution are best divided into two groups:

- the algorithms of making calculations;

- the algorithms of logic in decision-making.

The algorithms of making calculations ensure the use of computers to perform operational-tactical calculations for making a decision on the use of air defense weapons.

The algorithms of the logic of developing a decision are the most complex, since the logical function of man during the direct development of controlling action is the most difficult to formalize. The difficulties in formalizing this process are to a considerable degree explained by the complexity in the quantitative accounting of all factors affecting the decision-making. Here great significance is acquired on one hand by the formation of the logic circuits which determine the activity of man when making one decision or another, on the other hand - the separation of the decisive factors and their connections, which determine the selection of a solution.

The algorithms of transmitting information ensure the coding of report, the output of it from the computer into the communication channel and the output of a signal on the transmission of the reports.

CHAPTER 2

BASIC DATA FROM THE STATISTICAL DECISION THEORY IN CONNECTION WITH THE DETECTION OF SIGNALS

2.1. Reception of Signals as a Statistical Problem.

For the automation of the processes of controlling air defense forces and means, it is necessary to have comprehensive and continuous information on the coordinates and characteristics of air targets. This information is obtained, as a rule with the help of circular or sector scan radar. Radar target detection is based on the separation of radar signals reflected from targets against the background of noises which are random in nature. Noise distorts the signal, hindering the detection of the target and the determination of its parameters. The ability of radar to distinguish the useful signals against the background of internally-produced receiver noise is called observability.

Observability depends on the type of receiver, especially its terminating device (more precisely, on the nonlinear circuit of the receiver, beginning with the detector). The type of terminal equipment depends on the view of the observer who obtains the information.

The observer can be:

- a man (operator);
- a computer or a continuous-action automated machine;
- a discrete calculation computer - a digital computer.

Depending upon these factors the termination device of the radar can be:

- the indicator, if the observer is a man;
- the aerial-target automatic tracking equipment, if the observer is a continuous-action machine;
- the automatic data reduction equipment, if the recipient is the digital computer.

Each of the terminal equipment has elements which, together with the band-pass filter of the receiver, determine its ability to isolate the received signals against the background of noises. In the indicator such elements are the luminophor of the screen in conjunction with the scan, in continuous automatic tracking [AT] (AC) devices - a selection and lock-on system, in the automatic data reduction equipment - the preliminary selection unit (the preselector).

The terminal equipment, together with the problem of signal isolation, should solve the problem of radar data output to the observer. These data should be issued to the observer in a form acceptable for him, since none of the observers can directly use information in the form in which it is contained in the recieved signal. Thus, the indicator converts the radar signal to a form capable of actuating the senses of man.

Of all the five senses of man only sight and hearing can be used to perceive the obtained radar information. Consequently, indicators can be either visual or audible. The capabilities of visual indicators are considerably greater than audible indicators. For this reason visual indicators on cathode-ray tubes are the most widespread. The data on them are read off according to the scale of the screen on which the target coordinates are converted into linear blip displacements with the help of scanning.

A computer or a continuous-action actuating mechanism deals with the quantities represented in the form of electric currents or the voltages, angular or linear displacements of the elements in electromechanical systems which are analogs or models, and measured quantities included in the signal. Therefore, they were called analog, or simulating, computers. The terminal equipment connected with the simulator should issue information on target coordinates in the form of currents (voltages) and displacements.

Usually range data are issued in the form of voltages, and data on the angular position of the target - like the axial position of the antenna tracking the target.

The automatic data reduction [ADR] (АСД) equipment must feed the target data to the digital computer in the form of numbers (usually a binary code). Thus, it contains the elements for converting the measured quantities into numbers of binary coding.

The classification of terminal radar equipment, depending on type of observer, with subdivision into elements which perform the tasks of isolating the signal against the background of noises and converting the measurement data into a form acceptable for the recipient, is given in Table 2.1.

Table 2.1

Observer	Radar terminal equipment	Signal isolation elements	Coordinate-measuring elements
Operator	Indicator	Luminophor of the plan position indicator	Scale and scan
Continuous-action computer (or automated machine)	Automatic target-tracking device	Target lock-on and selection system	System for the output of data in the form of currents (voltages)
Digital computer	Automatic data reduction equipment	Preliminary selection unit (preselector)	Data output system with binary coding

The conversion of radar data is purely technical. It is necessary only so that the conversion will be qualitative, without reducing the accuracy of measurement inherent in the radar.

The problem of isolating a signal or detecting a target, in spite of the great distinction in the technical execution methods of terminal equipment and receivers as a whole, is described identically in the ideal case. However, each real device has these or other limitations which lead to decreased quality in observability.

The observability of signals is determined by the parameters of the signal and by properties of the noise, as well as by the subjective features of the observer and by his working conditions. The majority of these parameters bears a random, statistical nature.

It is known that noise is a random function of time which obeys the statistical laws of the theory of probabilities. A weak signal in interaction with noise also becomes a random quantity even in a case where it initially has a regular nature. In actuality the signal incoming at the receiver input is frequently also random quantity, even before the effect of noises, because of the "blinking" of the target, which is caused by the interference of signals reflected from its individual sections with a change in the orientation of the target.

In view of this, the problem of signal isolation in noises is statistical and can be analytically solved only by the methods of the theory of probabilities. This is all the more correct if the observer is a man.

Such operator qualities as physical fitness, attentiveness, and fatigue yield to no other calculation, except statistical.

2.2. Density of Probability Distribution of the Amplitude and Phase of Noise

The voltage of noises at the input of a radar receiver (antenna noise) has the most irregular structure (Fig. 2.1).

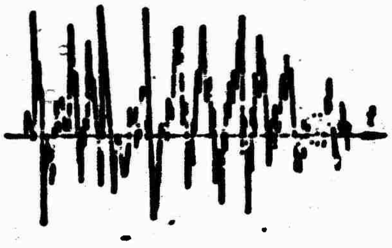


Fig. 2.1. Shape of the noise voltage at the receiver input.

The noise spectrum can be considered uniform up to the very high frequencies. Such noise is called "white" noise.

At the output of the linear part of the receiver the noise structure varies sharply: only the narrow band of frequencies Δf near resonance frequency f_0 in the band-pass filter of the linear frequency of the receiver (Fig. 2.2b) is eliminated from the entire wide spectrum (Fig. 2.2a). Noise oscillations at output of the band-pass filter can no longer be changed randomly. They assume the form of almost sinusoidal oscillations with a medium frequency of f_0, while the amplitude and the phase of these oscillations cannot rise or fall faster than the bandwidth of the filter allows.

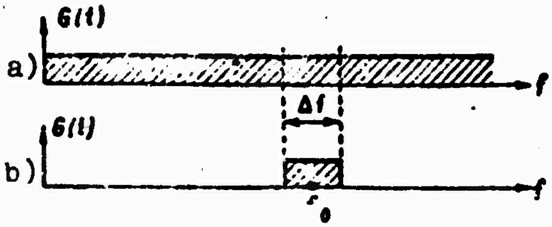


Fig. 2.2. The noise spectrum: a) at the input of the linear part of the receiver; b) at the output of linear part of the receiver.

Slow changes in the amplitude of noise oscillations are easy to explain, if the property of the process is connected with the parameters of a real physical system in which it is formed. If the system (oscillation circuit) is narrow-band, i.e., it has a high Q-factor, then it also has a slow response. As a result a considerable expenditure of energy is necessary to change the state of the system, and with limited power of the perturbations causing the oscillations it occurs slowly.

A narrow-band system is difficult to rapidly drive or stop, precisely as it is to change the phase of the oscillations. Thus, the amplitude (envelope) and phase of the high-frequency noise oscillations of frequency f_0 change comparatively slow.

Such a nature of the narrow-band noise process makes it possible to analytically present it in the form of almost sinusoidal oscillations with random and slowly changing amplitude and phase.

The equation of the process in this case can be written in the form

$$u(t) = U(t)\cos[\omega_0 t + \varphi(t)], \tag{2.1}$$

where U(t) and $\phi(t)$ - the amplitude and phase of oscillations (both random functions of time);

$\omega_0 = 2\pi f_0$ - the medium frequency of the oscillation spectrum.

If in expression (2.1) we find probability laws which the amplitude and phase of the oscillations obey, then the "noise" oscillation will thereby be fully determined. Quantity U(t) - the amplitude examined in the equation as a function of time - is called the envelope of the fluctuation (noise) process.

The concept of envelope is analogous here to the same concept which exists for amplitude-modulated oscillation, with only difference being that in the fluctuation process case the envelope, like fluctuation process itself, has a random nature.

Let us determine the probability density for quantities U and φ, assuming that process itself (its instantaneous values) obeys the normal distribution law.

Let us by means of the trigonometric expansion of equation (2.1) present fluctuation process in the form of the sum of two components (sinusoidal and consinusoidal):

$$x(t) = U(t)\cos[\omega_0 t + \varphi(t)] = U(t)\cos\omega_0 t\cos\varphi(t) - \\ - U(t)\sin\omega_0 t\sin\varphi(t).$$

We will designate:

$$U_1(t) = U(t)\cos\varphi(t); \\ U_2(t) = U(t)\sin\varphi(t),$$

then:

$$x_1(t) = U_1(t)\cos\omega_0 t; \\ x_2(t) = U_2(t)\sin\omega_0 t. \qquad (2.2)$$

Subsequently, to shorten writing let us present expression (2.2) in the following form:

$$x_1 = U_1\cos\omega_0 t; \\ x_2 = U_2\sin\omega_0 t. \qquad (2.3)$$

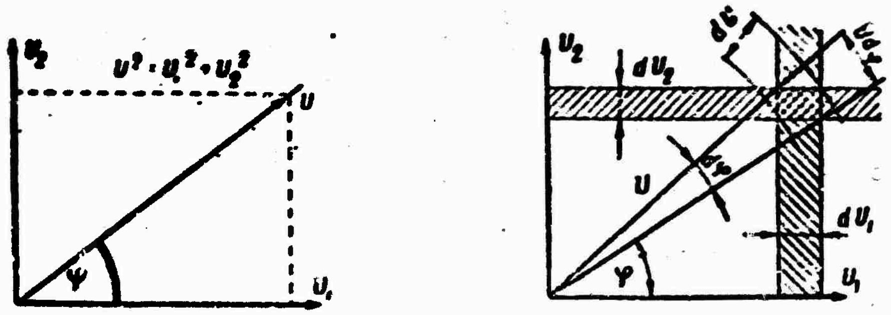
If fluctuation process (2.1) obeys the normal distribution law, then its components U_1 and U_2 also obey the normal distribution law. Consequently:

$$W(U_1) = \frac{1}{\sigma\sqrt{2\pi}}\exp\left[-\frac{U_1^2}{2\sigma^2}\right]; \\ W(U_2) = \frac{1}{\sigma\sqrt{2\pi}}\exp\left[-\frac{U_2^2}{2\sigma^2}\right], \qquad (2.4)$$

where $\sigma^2 = U^2_{\text{эфф}}$ - dispersion, i.e., the strength of the noise per resistance in 1 ohm.

These components are usually formed in a real physical system because of the overall effect of a large number of elementary consinusoidal and sinusoidal perturbations independent of each other, and they are independent random quantities. Quantity U is connected with its components by the relationship (Fig. 2.3a)

$$U = U_1^2 + U_2^2. \tag{2.5}$$



a) b)

Fig. 2.3. For the question of determining the density of probability distribution of the envelope and the phase of the noise at the input of linear part of the receiver: a) components of the process; b) method of determining the probabilities.

On the basis of the probability distribution laws (2.4) of quantities U_1 and U_2, we can find the probability that the end of vector U will get into the elementary sector with sides dU_1 and dU_2 (Fig. 2.3b).

This probability is equal (according to the probability multiplication theorem) to the product of the hit probabilities of each of the independent random quantities U_1 and U_2 into its elementary band dU_1, dU_2 and can be written in the form

$$W(U_1; U_2)\,dU_1 dU_2 = \frac{1}{2\pi\sigma^2}\exp\left[-\frac{U_1^2+U_2^2}{2\sigma^2}\right]dU_1 dU_2. \tag{2.6}$$

Quantity $W(U_1; U_2)$ is the combined (two-dimensional) density of probability distribution of quantities U_1 and U_2. In order to find the density of probability distribution of the envelope and the phase (U and ϕ), one should change from rectangular coordinates U_1 and U_2 to polar coordinates U and ϕ.

In this case it is necessary that the hit probability of U and ϕ into the element of an area in new coordinates, calculated the probability through the determined probability density $W(U, \phi)$, equals the hit probability into the corresponding (taking into account the functional connection between the old and new coordinates) elementary sector in the old coordinates. These probabilities should be equal, since between the old and the new coordinates there is a rigid functional connection. By taking into account (2.5) and (2.6), on the basis of the given considerations we can write:

$$W(U, \varphi)\, dU d\varphi = \frac{1}{2\pi\sigma^2} \exp\left[-\frac{U^2}{2\sigma^2}\right] U dU d\varphi. \qquad (2.7)$$

Hence the probability density

$$W(U, \varphi) = \frac{U}{2\pi\sigma^2} \exp\left[-\frac{U^2}{2\sigma^2}\right]. \qquad (2.8)$$

Quantity $W(U, \phi)$ is the combined density of probability distribution of random quantities U and ϕ - the envelope and the phase of the noise process. Quantity ϕ in clear form does not enter expression (2.8). This is explained by the fact (it will be examined more in detail below) that the probability density in this case does not depend on a concrete value of ϕ.

In order to explain which nature the probability densities of the envelope and the phase of the noise process have, it is necessary to derive expressions for each of these quantities individually.

It is known from the probability theory that at an assigned combined density of probability distribution of two random quantities, to get the density of one of them, it is necessary to sum up the probabilities of this quantity at all the possible values of the other. By using integration within the limits of all possible values of the phase from 0 to 2π, we will obtain the following expression for the density of probability distribution of the envelope of the noise process:

$$W(U)=\int_0^{2\pi} W(U,\varphi)\,d\varphi=\frac{U}{2\pi\sigma^2}\int_0^{2\pi} e^{-\frac{U^2}{2\sigma^2}}\,d\varphi= =\frac{U}{\sigma^2}\exp\left[-\frac{U^2}{2\sigma^2}\right]. \qquad (2.9)$$

Let us introduce dimensionless quantity x, which does not depend on the effective noise voltage:

$$x=\frac{U}{\sigma}\,(\sigma=U_{\text{эфф}}). \qquad (2.10)$$

On the basis of the rule of transition from one random variable (under the assigned law of its distribution) to another, we will have

$$W(x)\,dx=W(U)\,dU,$$

whence

$$W(x)=W(U)\frac{dU}{dx},\ \text{а}\ dx=\frac{1}{\sigma}\,dU. \qquad (2.11)$$

Thus, the expression for the density of probability distribution of the envelope can be reduced to the following form:

$$W(x)=\frac{U}{\sigma^2}\exp\left[-\frac{U^2}{2\sigma^2}\right]\frac{dU}{dx}=x\exp\left[-\frac{x^2}{2}\right]. \qquad (2.12)$$

The law of probability distribution described by expression (2.12) and shown in Fig. 2.4 is known by the name of the Rayleigh probability distribution, or the circular Gaussian distribution, law.

Thus, if the random process (noise) obeys the normal distribution law, then at the output of linear part of the receiver its envelope obeys the Rayleigh probability distribution law.

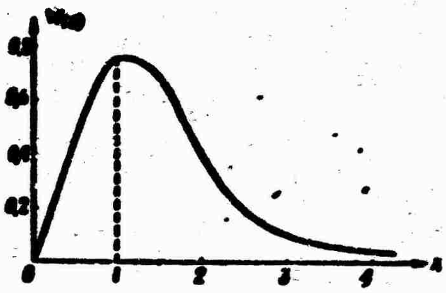


Fig. 2.4. The curve of the density of probability distribution of the envelope of the noise.

From the curve (Fig. 2.4) it is apparent that the maximum value of the density of probability distribution of the envelope takes place when x = 1 or when

$$U_{\max} = U_{\text{eff}} \qquad (2.13)$$

i.e., at a level numerically equal to the effective value of the noise voltage.

With an increase or decrease in this level, the density of probability distribution drops, approaching zero when $x \to 0$ and when $x \to \infty$.

Figure 2.5 gives the curve of the probability distribution for the envelope placed on a single diagram with the noise process image. Since the amplitude of noise in the positive and negative parts of the period is identical (MN = MP), the distribution curve of the envelope is given in the area of only positive values of x.

Since the envelope is a random function of time, in accordance with which the amplitude of noise changes, on the basis of the distribution curve in (Fig. 2.4) and equality (2.13) it can be asserted that among the different values of the noise oscillation

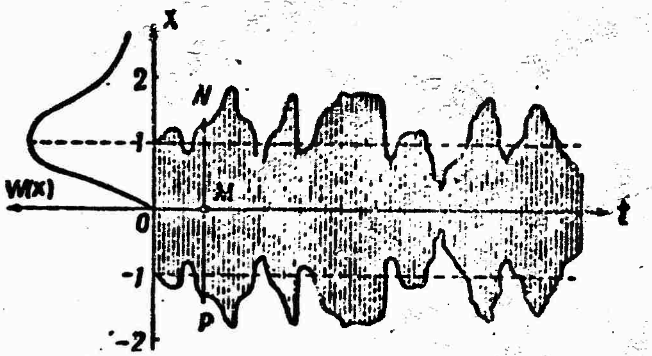


Fig. 2.5. The curve of the distribution density of the envelope in conjunction with the image of the noise process.

one should most frequently expect the appearance of the oscillations with an amplitude equal or close to the effective value of the voltage. Such amplitude values, as they say, are the most probable. Other amplitude values will be encountered relatively seldom with the medium frequency being characterized by the appropriate values of the density probability distribution for the envelope.

Let us now determine the phase of the noise process.

By integrating expression (2.9) within the limits of all possible values of the envelope (from 0 to ∞), we obtain the following expression for the density of probability distribution for the phase of the noise process:

$$W(\varphi)=\int_0^\infty W(U,\varphi)\,dU=\frac{1}{2\pi}\int_0^\infty \frac{U}{\sigma^2}e^{-\frac{U^2}{2\sigma^2}}\,dU=\frac{1}{2\pi}. \tag{2.14}$$

From expression (2.14) it is apparent that the probability density of the phase is constant throughout the entire range of its possible values from 0 to 2π. This, in particular, means that if we isolate on the time axis some arbitrary interval Δt of observation time (let us assume it equal to the average period of noise oscillation $T_0 = \frac{2\pi}{\omega_0}$), then the noise oscillation can with equal probability intersect this interval at any point of it.

2.3. Mathematical Description of the Interaction of a High-frequency Signal and Noise

We will now find the law of probability distribution for the envelope of a process where the signal acts jointly with the noise in the receiver.

A sequence of pulses whose amplitude is a random quantity and can be varied over the wide limits from zero to very high values is obtained as a result of superimposing noises on the pulses of a signal of constant amplitude.

At the same time, when there is no signal, pure noise can cause large overshoots which in no way differ from the pulses of the signal.

The interaction of the signal and noise can occur also in such a way that at the beginning of the signal pulse they are put in one phase together, and at the end of the pulse - in another. As a result of this the obtained total pulse can be displaced with time. Since the range to the target is judged by the position of pulses, such an influence of noises on the signal leads to errors in measurement.

Thus, as a result of superimposing noises on the signal the suppression of a useful signal, the appearance of a spurious signal and errors in measurement are possible.

Let us present the equation for the signal in the form

$$u_c(t) = U_c \cos\omega_0 t,$$

and the noise oscillation at the output of the linear part of the receiver in the form of almost sinusoidal oscillations with a random amplitude and phase:

$$u_{\text{ш}}(t) = U_{\text{ш}} \cos[\omega_0 t + \varphi_{\text{ш}}(t)]. \tag{2.15}$$

We will assume that noise oscillation obeys the normal distribution law with the average value equal to zero. Let us expand the envelope of noise oscillations $U_{\text{ш}}(t)$ by two components in such a way that one of them will coincide with the signal in phase (the phase component), and the other will be out of phase by 90° (out-of-phase component).

As a result we will obtain

$$u_{\text{ш}}(t) = U_{1\text{ш}} \cos \omega_0 t + U_{2\text{ш}} \sin \omega_0 t,$$

where

$$U_{1\text{ш}} = U_{\text{ш}}(t) \cos \varphi(t);$$
$$U_{2\text{ш}} = \dot{U}_{\text{ш}}(t) \sin \varphi(t);$$
$$U_{\text{ш}}(t) = \sqrt{U_{1\text{ш}}^2 + U_{2\text{ш}}^2}.$$

The vector diagram U_c of the signal, together with noise $U_{1\text{ш}}$ and $U_{2\text{ш}}$, is shown in Fig. 2.6a. $U_{c\text{ш}}$ represents the amplitude (or envelope) of the resulting oscillation of the signal and noise, but ϕ is its phase. In the presence of the signal phase ϕ of the total voltage differs from phase $\phi_{\text{ш}}$ of pure noise.

Let us introduce the new variables $U_1 = U_{1\text{ш}} + U_c$ and $U_2 = U_{2\text{ш}}$.

The distribution laws for each of these variables can be written in the form:

$$W(U_1) = \frac{1}{\sigma_{\text{ш}}\sqrt{2\pi}} \exp\left[-\frac{(U_1 - U_c)^2}{2\sigma_{\text{ш}}^2}\right],$$
$$W(U_2) = \frac{1}{\sigma_{\text{ш}}\sqrt{2\pi}} \exp\left[-\frac{U_2^2}{2\sigma_{\text{ш}}^2}\right], \tag{2.16}$$

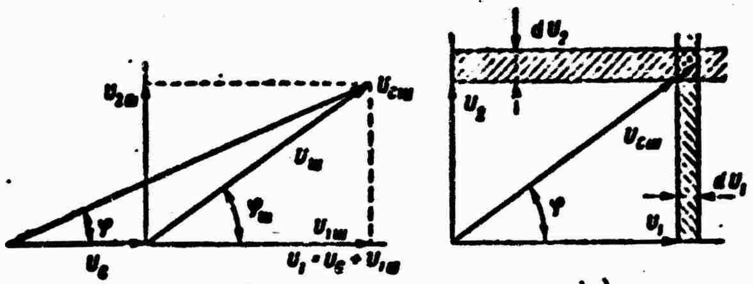
where $\sigma_{\text{ш}} = U_{\text{эфф}\cdot\text{ш}}$ is the effective voltage of the noise.

On the basis of formula (2.16) let us find the expression for determining the probability that quantity U_1 will assume a

value lying within the limits of U_1 and $U_1 + dU_1$, and quantity U_2 - simultaneously within the limits between U_2 and $U_2 + dU_2$. Since quantities U_1 and U_2 are independent, this probability according to the probability multiplication theorem is equal to the product

$$W(U_1)dU_1W(U_2)dU_2=\frac{1}{2\pi\sigma_ш^2}\exp\left[\frac{(U_1-U_c)^2+U_2^2}{2\sigma_ш^2}\right]dU_1dU_2 \qquad (2.17)$$

The obtained expression determines the hit probability of the end of the vector for the resulting voltage $U_{cш}$ into elementary sector dU_1 and dU_2 (Fig. 2.6b).



a) b)

Fig. 2.6. For the problem of determining the density of probability distribution of the "signal pulse noise" mixture: a) the voltage components; b) the method for determining the probabilities.

Expression (2.17) is consistent with expression (2.6) for noise without the signal, with the exception of the exponent at right side of expression (2.17). The exponent in expression (2.17) can on the basis of vector diagrams in (Fig. 2.6) be represented in the form

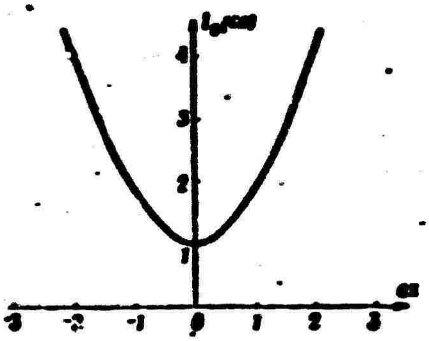
$$\frac{(U_1-U_c)^2+U_2^2}{2\sigma_ш^2}=\frac{U_{cш}^2+U_c^2-2U_{cш}U_c\cos\varphi}{2\sigma_ш^2},$$

then expression (2.17) will assume the following form:

$$W(U_1;U_2)=\frac{1}{2\pi\sigma_ш^2}\exp\left[-\frac{U_{cш}^2+U_c^2-2U_{cш}U_c\cos\varphi}{2\sigma_ш^2}\right]. \qquad (2.18)$$

For obtaining the probability density for the envelope of noise plus the signal ($U_{cш}$) it is necessary to pass from the

Fig. 2.7. The shape of the function $I_0(ax)$.



$$W(x) = x\exp\left[-\frac{x^2+a^2}{2}\right] I_0(ax). \qquad (2.24)$$

Figure 2.8 gives the curves of the density of probability distribution, which were constructed with respect to formula (2.24). The law of probability distribution of the envelope of the signal plus noise, which is described by expression (2.24), is called the generalized Rayleigh law, or the Rice law. In a particular case when $a = 0$ (when there is no signal), we obtain the common Rayleigh distribution (2.12):

$$W(x) = xe^{-\frac{x^2}{2}} = x\exp\left[-\frac{x^2}{2}\right]. \qquad (2.25)$$

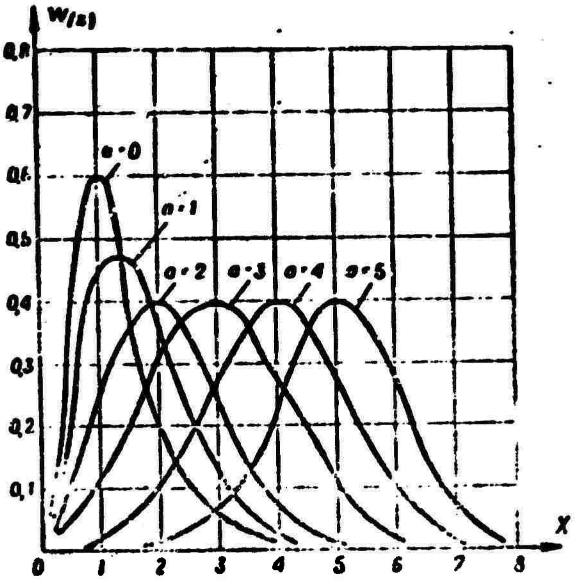


Fig. 2.8. The density of probability distribution for amplitude at different values of a.

For the complete characteristics of the behavior of the signal in the noises, of interest also is the density of distribution of phases $W(\phi)$, which is determined by averaging $W(x, \phi)$ from expression (2.21) by all values of x from zero to infinity:

$$W(\varphi)=\frac{1}{2\pi}\exp\left[-\frac{a^2}{2}\right]+\frac{a\cos\varphi}{\sqrt{2\pi}}F(a\cos\varphi)\exp\times\\ \times\left[-\frac{a^2\sin^2\varphi}{2}\right]. \qquad (2.26)$$

where $F(z) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{z} e^{-\frac{t^2}{2}}dt$ - the Laplace function ($z = a\cos\phi$).

This distribution is given in Fig. 2.9.

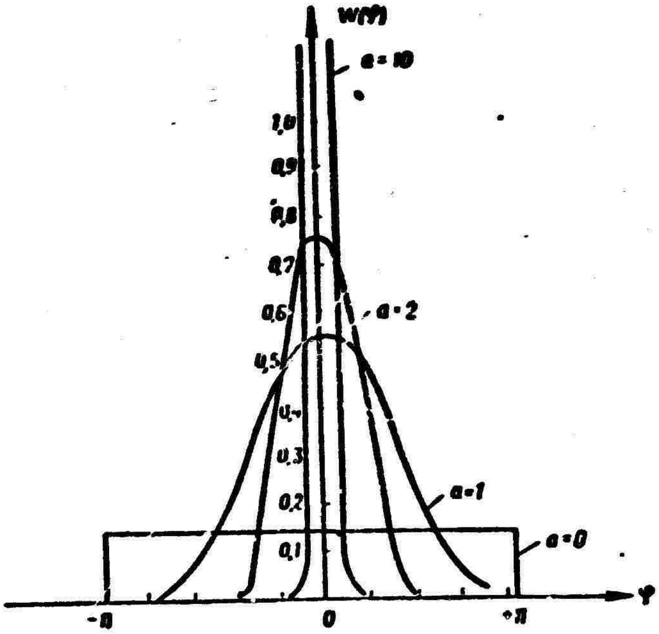


Fig. 2.9. The density of probability distribution for the phase at different values of a.

By analyzing Figs. 2.8 and 2.9, we notice that to the degree the intensity of the signal increases, the law of the distribution density of envelope x approaches a symmetrical, normal law.

The law of distribution density of the phase with increased a also becomes normal, whereupon the most probable values of the oscillation phase concentrate close to $\phi = 0$. With a very powerful signal the overall oscillation phase barely differs from the phase of a pure signal ($\phi = 0$), and is therefore stable.

This can be explained by the fact that little noise, combining with the signal, cannot change the phase of the signal significantly.

2.4. Probability Characteristics of the Quality of Radar Detection

The quality of work of an automated system in processing radar information primarily depends on the best way of performing each processing operation. The optimum algorithms of processing are synthesized in accordance with some preassigned criteria selected from the totality of the problems being solved and the specific requirements which are placed on the solutions. Since the processed radar information is random in nature and can be only described statistically, the criteria of the optimality of the processing algorithms should also be statistical. In connection with this, as the general theoretical basis for synthesizing the processing algorithms it is advantageous to use the theory of statistical solutions, which gives the general approach to selecting the best rules in processing random signals. The basic advantage of this theory is the fact that it unites the whole variety of optimality criteria utilized in statistics, whereupon these criteria ensue from the general theory as exceptional cases.

Let us examine the basic positions of the general theory of solutions and some of its results in connection with the problem of processing radar signals obtained by radar sites. After processing the obtained information a decision on the presence or absence of a useful signal must be made. By a useful signal is meant one pulse or a batch of pulses reflected from the target. This decision is made objectively under two incompatible conditions which characterize the status of the target:

$$\left.\begin{array}{l} 1)\ \text{"Target"} - H_1; \\ 2)\ \text{"No Target"} - H_0. \end{array}\right\} \qquad (2.27)$$

The assigned task pertains to the so-called two-alternative situation, where one of two events occurs: "target" or "no target." In this case some observations, whose combination in mathematical statistics is called sampling, are made. The nature of the observations is such that it is not possible with reliability to establish by them (because of the presence of interferences and noises in the signal) which of the two events actually took place. Only the probability connections between the events and the observations are known.

Problem is to select one of two hypotheses by the results of the observations: hypothesis H_1^* (the hypotheses, i.e., the presumable solutions, we will designate by the asterisk) of the fact that "there is a target," or hypothesis H_0^* of the fact that "there is no target."

The hypotheses mutually eliminate each other. Selecting one of them is also called the statistical solution. Below, the criteria and rules of solution will be discussed.

By criterion is meant some general conditions which should satisfy the selection of a hypothesis, i.e., the solution. These conditions frequently have an extremal nature, i.e., it is necessary that the solution will minimize or will maximize these or other quantities.

By rule is meant the description of the specific procedure which must be accomplished in order to obtain a solution which is satisfactory to this criterion. In the simplest case the criterion can consist of the fact that of the two hypotheses the more probable one is selected. In other words, that hypothesis which is correct with the greater probability is selected.

With the two possible events and correspondingly the two hypotheses, the following four cases decision-making are possible (Table 2.2).

Table 2.2

Actual event	Selected hypothesis	Decision
Target - H_1	H_1^* (target)	Correct (correct detection)
Target - H_1	H_0^* (no target)	Erroneous (target penetration)
No target - H_0	H_1^* (target)	Erroneous (false alarm)
No target - H_0	H_0^* (no target)	Correct (correct nondetection)

As is evident from the table, erroneous and correct decisions are possible. The wrong decisions arise in two cases (the absence and the presence of a target) and are correspondingly called: a false alarm, when the decision is that there is a target, while in actuality there is no target; penetration of a target, when the decision is that there is no target, while in actuality a target exists. The two forms of correct decisions also correspond to the presence and the absence of a target and are correspondingly called: a correct detection, when the decision is that there is a target, and in fact there is one: a correct nondetection, when the decision is that there is no target while there is actually no target.

Each of these cases is characterized by an unconditional probability, which in accordance with the probability multiplication theorem can be presented in the following form:

- for the correct detection

$$p(H_1^*, H_1) = p(H_1)\,p(H_1^*/H_1);$$

- for the false alarm

$$p(H_1^*, H_0) = p(H_0)\,p(H_1^*/H_0); \tag{2.28}$$

- for the penetration of a target

$$p(H_0^*, H_1) = p(H_1)\,p(H_0^*/H_1);$$

- for the correct nondetection

$$p(H_0^*, H_0) = p(H_0)\,p(H_0^*/H_0),$$

where $p(H_1)$ and $p(H_0)$ are the *a priori* probabilities of the presence and the absence of a target (signal); $p(H_1^*/H_1)$, $p(H_1^*/H_0)$, $p(H_0^*/H_1)$ and $p(H_0^*/H_0)$ - the corresponding *a posteriori* probabilities, i.e., the conditional probabilities calculated under the assumption of the actual presence or the absence of a signal.

As can be seen from relationships (2.28), the unconditional probabilities can be determined, if the *a priori* and the *a posteriori* probabilities are known. The *a priori* probabilities (below, to shorten writing we will designate $p(H_1) - P_1$, $p(H_0) = P_0$) can be determined using the known preliminary data on target location. If this is impossible, then there appears the so-called *a priori* difficulty, which frequently occurs in radar. In this case it is possible to consider the absence and the presence of a signal to be equally probable ($P_1 = P_0 = \frac{1}{2}$). To calculate the *a posteriori* probabilities of threshold acquisition systems, it is necessary to know the functions of the density of conditional probability distribution of the investigated process.

Let us examine the functions of the density of probability distribution of noise $W_{ш}(x)$ and the "signal plus noise" mixture $W_{сш}(x)$, where x is the argument of the distribution function (2.20). The shape of the curves for the indicated functions is given in Fig. 2.8. We will select two of them when a = 0 and a ≠ 0 (Fig. 2.10).

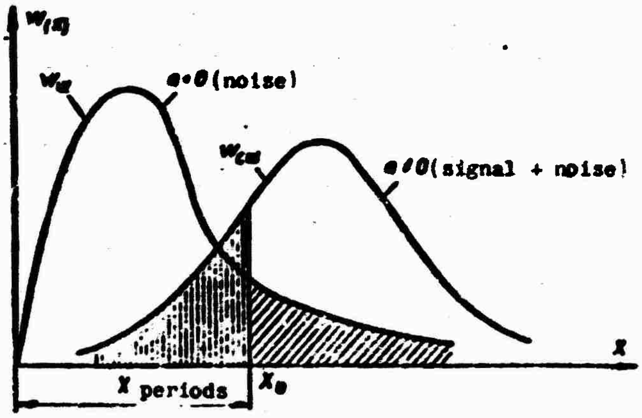


Fig. 2.10. For determining the probabilities of a false alarm and target admission.

If we select some relative value of accumulation threshold x_0 ($x_0 = \frac{U_{сш0}}{\sigma}$, where $U_{сш0}$ is the absolute value of the threshold), then the *a posteriori* probabilities which characterize the quality of radar detection can be found from the following expressions.

The probability $P_{по}$ of correct detection is the area bounded by the section of curve $W_{сш}(x)$ and by the axis of abscissa situated more to the right of threshold line x_0, i.e.,

$$P_{по}=\int_{x_0}^{\infty} W_{сш}(x)\,dx. \tag{2.29}$$

The probability $P_{пц}$ of target penetration is the area located under curve $W_{сш}(x)$ situated more to the left of threshold line x_0, i.e.,

$$P_{пц}=\int_{0}^{x_0} W_{сш}(x)\,dx. \tag{2.30}$$

The probability $P_{лт}$ of a false alarm is characterized by the area situated under the section of curve $W_ш(x)$ situated more to the right of threshold x_0, i.e.,

$$P_{лт} = \int_{x_0}^{\infty} W_{ш}(x)\,dx. \qquad (2.31)$$

And, finally, the probability $P_{пн}$ of correct nondetection is characterized by the area situated under the section of curve $W_{ш}(x)$ more to the left of the threshold of limitation, i.e.,

$$P_{пн} = \int_{0}^{x_0} W_{ш}(x)\,dx. \qquad (2.32)$$

With consideration that the area bounded by curve $W_{ш}(x)$ or $W_{сш}(x)$ equals unity, since it characterizes an authentic event, these relationships occur:

$$\begin{gathered} P_{по} + P_{пп} = 1; \\ P_{лт} + P_{пн} = 1. \end{gathered} \qquad (2.33)$$

As is evident, of the four parameters which characterize the quality of detecting signals against the background of noises, only two are independent. Any pair of independent quantities can be taken as the initial characteristic of quality indices for the operation of signal detection equipment.

Probability $P_{по}$ of correct detection and probability $P_{лт}$ of false alarm are assigned most frequently during analysis.

2.5. Selection of the Criteria for the Quality of $P_{по}$ and $P_{лт}$ Detection

Selecting the optimum quality indices of $P_{по}$ and $P_{лт}$ detection is a contradictory problem which can be solved only on the basis of intelligent compromise. Actually, the high probability $P_{по}$ of correct detection and low probability $P_{лт}$ of a false alarm are natural for the correct operation of detection equipment. However, it is impossible to simultaneously satisfy these two requirements, since to increase $P_{по}$ (2.29) it is necessary to diminish the threshold value, while to reduce $P_{лт}$ (2.31) the threshold must be increased. Thus, it is necessary to go to an intelligent compromise, selecting the optimum regime for detection equipment from the viewpoint of the totality of the possible processing conditions.

Naturally, the solution found in this case will not necessarily be the best from the viewpoint of any particular condition, for example, the maximum of the probability of correct detection or the minimum of a false alarm. The decision made should be optimum on the average for the different operating conditions of the detection equipment, with consideration of the statistics of distribution, the possible detection conditions and the importance of one or another situation. The average risk method has found the widest application. Let us examine the essence of the method.

The entire totality of the radar detection situation consists of correct detection, target penetration, false alarm and correct nondetection. From the quantitative side each of these situations is characterized by its unconditional probability (2.28). The concept of price is introduced for every situation. The significance of price is placed into conformity with the importance of permissible error in one or another situation. Thus, for instance, in certain cases the cost for target penetration can be considerably higher than for a false alarm. Actually, the result of target penetration can be the destruction of the defended object, while a false alarm is usually connected with considerable material expenditures.

Correct detection and correct nondetection have a zero value of cost, i.e., error-free decisions have a zero cost. In this case it is important so that the price of an error in any incorrect decision will be greater than price of a correct one and so that its greater cost will correspond to the large error. Then the average risk q_1 of each situation, or the amount of cost "paid off" in each case, is proportional to appearance probability P_i and cost r_i of the error in this situation, i.e.,

$$q_i = r_i P_i. \tag{2.34}$$

When $r_i = 0$ (it corresponds to the error-free situation), and also when $P_i = 0$ (it corresponds to the improbable situation), $q_i = 0$. In essence, expression (2.34) represents the loss paid off for the risk of making a decision in the i-th situation.

Average risk R of totality of n situations is calculated by the rules of mathematical expectation and for discrete cases is equal to the sum of the losses obtained from the individual situations:

$$R = \sum_{i=1}^{n} q_i = \sum_{i=1}^{n} r_i P_i.$$

During analysis of the detection process it is sufficient to set:

- the cost of a false alarm error

$$r_{\text{лт}} = r(H_1^*, H_0);$$

- the cost of a target penetration error

$$r_{\text{пц}} = r(H_0^*, H_1),$$

since the cost of errors in correct detection and correct nondetection equals zero. Then, only terms which characterize erroneous detection will be present in the expression for the average risk:

$$R = r_{\text{лт}} P(H_1^*, H_0) + r_{\text{пц}} P(H_0^*, H_1).$$

This equality can be rewritten in the form

$$R = r_{\text{лт}} P_{\text{лт}} P_0 + r_{\text{пц}} P_{\text{пц}} P_1. \tag{2.35}$$

When the different processing systems are being compared, preference must be given to that system for which the average risk is less. Consequently, the optimum conditions and parameters of the detection equipment should be found in minimum criterion of the average risk, i.e., from the minimum condition of expression (2.35). Specifically, if we consider that the cost for the situation of a false alarm and target penetration is identical and equal to unity ($r_{\text{лт}} = r_{\text{пц}} = 1$), then the average risk will be equal to the sum of the probabilities of errors in detection:

$$R = P_{\text{лт}} P_0 + P_{\text{пц}} P_1. \qquad (2.36)$$

The minimum of this probability is called the ideal observer criterion. It is evident that the criterion of the minimization of average risk is more common in comparison with the ideal observer criterion, since it considers distinction in the cost of errors in the false alarm and in target penetration.

By substituting $P_{\text{пц}} = 1 - P_{\text{по}}$ into expression (2.35), we will obtain

$$R = r_{\text{пц}} P_1 - r_{\text{пц}} P_{\text{по}} P_1 + r_{\text{лт}} P_{\text{лт}} P_0.$$

By grouping the latter two terms and carrying $r_{\text{пц}} P_1$ beyond the bracket, we will obtain

$$R = r_{\text{пц}} P_1 - r_{\text{пц}} P_1 [P_{\text{по}} - l_0 P_{\text{лт}}], \qquad (2.37)$$

where

$$l_0 = \frac{r_{\text{лт}} P_0}{r_{\text{пц}} P_1}.$$

Since the first term is positive, the minimum criterion of the average risk is reduced to finding the maximum of the difference:

$$P_{\text{по}} - l_0 P_{\text{лт}} \qquad (2.38)$$

This condition requires such an increase in probability $P_{\text{по}}$ of correct direction and a decrease in probability $P_{\text{лт}}$ of a false alarm by which difference $P_{\text{по}} - l_0 P_{\text{лт}}$ increases. Multiplier l_0 is called the weight factor. It depends on the value of the *a priori* probabilities of the presence or the absence of a target in the investigated sector of space and the cost of an error of every kind. Let us assume that all of these values are assigned. If the two information processing systems are compared for the purpose of detecting a useful signal against a background of noises under identical conditions (identical l_0), then the best of them will be that one for which

$$P_{\text{по opt}} - l_0 P_{\text{лт opt}} \geqslant (P'_{\text{по}} - l_0 P_{\text{лт}}).$$

Then,

$$P_{\text{по opt}} \geqslant P'_{\text{по}} + l_0 (P_{\text{лт opt}} - P_{\text{лт}}),$$

and under the condition that $P_{\text{лт opt}} = P_{\text{лт}}$,

$$P_{\text{по opt}} \geqslant P'_{\text{по}}.$$

This means that with a fixed probability $P_{\text{лт}}$ of a false alarm the optimum system will be that one for which the probability $P_{\text{по}}$ of correct detection will be greatest. This condition is called the Neumann-Pearson criterion by the name of the known mathematicians, specialists in the field of mathematical statistics.

To obtain concrete quantities for $P_{\text{по}}$ and $P_{\text{лт}}$ it is necessary to assign the magnitude of the average risk R, as well as value l_0 and the amounts $r_{\text{пц}}$ and $r_{\text{лт}}$ of the cost of errors. In this case from expression (2.37) we find

$$P_{\text{по}} - l_0 P_{\text{лт}} = \frac{r_{\text{пц}} P_1 - R}{r_{\text{пц}} P_1}.$$

In radar the values for *a priori* probabilities P_0 and P_1, as a rule, are unknown. Their definition comprises the so-called *a priori* difficulty, since in the majority of cases there is no possibility of determining the probability of target appearance (ship, aircraft) in the detection area of a station at one or another instant. It is possible with identical success to speak in favor of one hypothesis or another, i.e., in favor of the presence or the absence of a target. Therefore, it is possible to consider that both hypotheses are equally probable, i.e., $P_0 = P_1 = 0.5$.

Sometimes it is also difficult to evaluate the cost for one situation or another. Actually, how does one evaluate the losses associated with a false alarm and target penetration? How does one quantitatively evaluate the results of panic among population caused by a false alarm or by blunting the alertness of personnel in the defense system as a result of a false alarm? Finally, in what kind of units are human sacrifices and destruction caused by the penetration of a target to the object measured? In a number of cases the lack of precise data on the cost of losses leads to the necessity for assuming $r_{\text{пц}} = r_{\text{дт}} = 1$.

Under these conditions $l_0 = 1$, and then

$$P_{\text{по}} - P_{\text{лт}} = \frac{0{,}5 - R}{0{,}5}.$$

This relationship can be used for rough estimates.

The examined criteria were accepted without any grounds. In connection with this, their correctness can be doubted, especially since it is always possible to expect that there are other optimum methods which use the information available in the input signal more fully and will yield the best results. Nevertheless, at the present the most common criterion is considered the minimum criterion of the average risk of losses (2.38).

By substituting $P_{\text{по}}$ and $P_{\text{лт}}$ from expressions (2.29) and (2.31) and taking into account that the integration for $P_{\text{по}}$ and $P_{\text{лт}}$ is conducted within identical limits, we will obtain

$$\int_{x_0}^{\infty} [W_{\text{сш}}(x) - l_0 W_{\text{ш}}(x)]\, dx = \max.$$

In order that the integral will be maximum, it is necessary to attribute to the domain of integration only those values of x for which integrand is positive, i.e.,

$$W_{\text{сш}}(x) - l_0 W_{\text{ш}}(x) > 0, \qquad (2.39)$$

or

$$\frac{W_{\text{сш}}(x)}{W_{\text{ш}}(x)} = \lambda > l_0.$$

Quantity λ is called the probability ratio, or the probability coefficient.

Let us note that probabilities $P_{\text{по}}$ and $P_{\text{лт}}$, in accordance with formulas (2.29) and (2.31), depend on threshold x_0. Hence it follows that $P_{\text{по}}$ and $P_{\text{лт}}$ are also interdependent. The dependence between the probabilities of correct detection and false alarm are frequently called the performance characteristics of the detector (receiver performance - RP), which is constructed in coordinates $P_{\text{по}}$ and $P_{\text{лт}}$ and characterizes the probability ratio. Since derivative of integral is equal in terms of lower limit to the integrand taken with a reverse sign from formulas (2.29) and (2.31), it follows that

$$W_{\text{сш}}(x_0) = -\frac{dP_{\text{по}}}{dx_0} \text{ и } W_{\text{ш}}(x_0) = -\frac{dP_{\text{лт}}}{dx_0};$$

$$\left(\frac{dP_{\text{по}}}{dP_{\text{лт}}}\right)_{x=x_0} = \frac{W_{\text{сш}}(x_0)}{W_{\text{ш}}(x_0)} = \lambda_0,$$

where λ_0 is the threshold value of the probability coefficients.

Thus, at any point of the characteristics of the detector the tangent of the slope angle of the tangent is numerically equal to the value of the threshold probability coefficient.

Performance passes through two characteristic points: ($P_{по} = 0$; $P_{лт} = 0$) when $\lambda_0 = \infty$ and ($P_{по} = 1$; $P_{лт} = 1$) when $\lambda_0 = 0$. In the first case the magnitude of threshold λ_0 selected infinitely large, and neither the signal, nor pure noise can exceed it. In the second case there is no threshold, and any signal can be accepted as useful.

If we return to formula (2.39), it is important to note that optimum detection is reduced to the calculation of probability ratio λ and to comparing it with threshold l_0. Thus, optimum detection also includes the operation of threshold comparison.

To calculate the probability ratio it is necessary to arrange the expressions of the distribution functions of the accepted "signal plus noise" mixture and the pure noise, as well as the sampling realization values of the accepted signal. The structure of signal processing for detection decision-making is reduced to a calculation operation in accordance with formula (2.39).

This formula can be examined as the initial expression for composing the algorithm of calculation which must be performed over the output signal realizations in order to detect a reflected pulse. Detector itself should be nothing else but a specialized-computer which performs calculations in accordance with the indicated formula.

CHAPTER 3

INITIAL PROCESSING OF RADAR DATA

3.1. Determination and Complement of the Initial Processing of Radar Data

Radar is one of the basic means of obtaining information on aerial targets. Radar network forms a radar system which provides information to air defense forces. This information is the result of radar signal analysis (the signals of a receiving circuit; signals which characterize location of an antenna system; time signals).

The process of obtaining information about objects found within radar visibility ranges is called radar data processing. Such processing makes it possible to obtain the data on target coordinates, its trajectory parameters (speed, acceleration, course angle), location time and others. Let us agree to call totality of information on the target the blip. The complement of the blips, besides the above-indicated data, can include information on target number, what government it belongs to, type, importance, authenticity, etc. The formed blip can subsequently be represented partially or completely by different kinds of signals.

For example, on plan-position indicators the blip is represented in the form of a luminous dot, and in automatic data processing machines - in the form of a binary number, voltage or current values.

The signals which will bear the information the observer (the operator) needs are called useful signals. As a rule, interferences are superimposed on useful signals. Because of these interferences the target information is distorted. In connection with this, during processing there arise problems in isolating the useful signals and obtaining necessary information under interference conditions. This compels one to search for data processing methods with whose aid the ill effect of interferences can be reduced to a minimum. However, it is impossible to get completely rid of them,on account of which the results of the processing always contain some errors.

Data processing is based on the existence of distinctions between the useful signal and the interference. Frequently, such a distinction is detected in the regularities of random signals coming in for processing. The entire process of radar data processing can be divided into three stages: initial, secondary and tertiary processing.

At the initial radar-data processing stage the target is detected and its coordinates are determined. Here the signals of the radar receiving circuit, antenna angular-coordinate sensors and time signals (synchronizing radar pulses are used). Initial processing is accomplished by one or several adjacent range sweeps. In principle these signals are sufficient to detect a target and to determine its coordinates. In connection with this, the initial processing of radar data (in connection with circular scan radar) is the processing of data for one radar scanning period.

The complement of the initial radar-data processing is: the detection of a useful signal in noise; the determination of target coordinates; coding the coordinates of the detected target and the initial numbering of targets.

3.2 Structural Diagram for the Optimal Detection of a Single Pulse Signal

The initial processing of radar data begins with the detection of an useful signal in noises. To explain just how the resolver circuit should look we will first examine the simplest case - detecting a single pulse signal. In this case the resolver circuit, as follows from the previous chapter, should be selected on the basis of the general criteria which define the detection problem.

The minimum criterion of average risk can be accepted as one of them. Below we will call the resolver which operates on the basis of the use of this criterion the optimal detection device, since it determines the best outcome when solving the problems of detection. The minimum criterion of average risk leads to comparing the likehood ratio with threshold l_0.

Let us assume that only amplitude value x(t) of the signal at any chosen instant t_1 is analyzed, i.e., the decision on the presence or the absence of a signal is accepted on the basis of one-dimensional sampling. The conditional densities of probability distribution for the envelope $W_{\text{сш}}(x)$ of the "signal plus noise" mixture and the envelope $W_{\text{ш}}(x)$ of noise voltage for the individual pulses (one sampling) are taken as known.

Let us find the algorithm for the optimal detection of a single pulse signal in narrow-band noise, using likelihood ratio $\lambda(x)$:

$$\lambda(x) = \frac{W_{\text{сш}}(x)}{W_{\text{ш}}(x)} = \frac{x\exp\left[-\frac{x^2+a^2}{2}\right] I_0(ax)}{x\exp\left[-\frac{x^2}{2}\right]} = \exp\left[-\frac{a^2}{2}\right] I_0(ax). \qquad (3.1)$$

According to the condition of the minimum criterion of average risk, the likelihood ratio for making the correct decision should be more than (or equal to) threshold value l_0, i.e.

$$\lambda(x) = \exp\left[-\frac{a^2}{2}\right] I_0(ax) > l_0 \qquad (3.2)$$

This relationship should be examined as the algorithm of the calculations made above sampling values of x(t) in the circuit of the optimum detector, which in essence is a computer. For the sake of simplicity in the computer let us proceed to the logarithmic threshold value, i.e., to quantity ln l_0. Let us note that due to the monotony of the logarithmic function, inequality (3.2) will not be violated; only the numerical values of the right and left sides will change:

$$\ln \lambda(x) = -\frac{a^2}{2} + \ln I_0(ax) \geqslant \ln l_0$$

Replacement of $\lambda(x)$ by ln $\lambda(x)$ requires the corresponding replacement of threshold l_0 by ln l_0. As a result we obtain resultant expression for the algorithm of the detection of a single pulse signal:

$$\ln I_0(ax) \geqslant \ln l_0 + \frac{a^2}{2}. \qquad (3.3)$$

The block diagram presented in Fig. 3.1 corresponds to this algorithm.

Expression (3.3) makes it possible to explain the structure of the resolver. After the nonlinear transformation of the envelope x(t) of the single pulse signal with the aid of equipment with characteristic $z = \ln I_0(ax)$, the obtained voltage should be compared with the level of limitation $l_0' = \ln l_0 + \frac{a^2}{2}$ created in the threshold device. Such a device is a detector with a characteristic which satisfies the condition

$$z = \ln I_0(ax). \qquad (3.4)$$

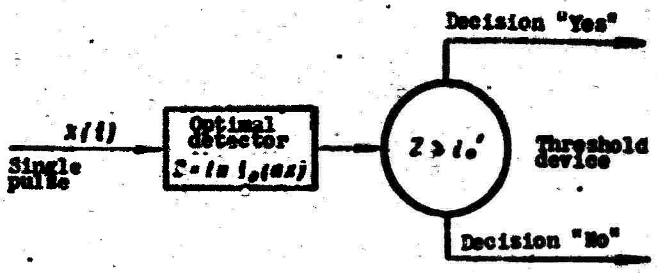


Fig. 3.1. Block-diagram of the resolver for detecting the single pulse signal.

At the same time the following approximate representation of function z exists:

$$\ln I_0(ax) \approx \begin{cases} \frac{(ax)^2}{4} \text{ when } ax \ll 1; \\ ax \text{ when } ax \gg 1. \end{cases} \qquad (3.5)$$

Case $ax \ll 1$ corresponds to a weak signal.

Case $ax \gg 1$ corresponds to a strong signal.

Thus, the optimal characteristic of the detector should be quadratic for signals which are weak in comparison with the noise and linear - for the remainder (Fig. 3.2).

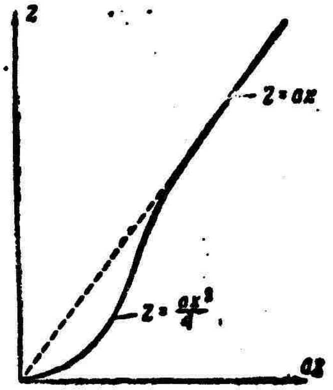


Fig. 3.2. The optimal characteristic of the detector.

How does one determine the magnitude of the threshold value for the likelihood ratio l_0 for example using the Neumann-Pearson criterion?

To do this, on the strength of assigned probability $P_{лт}$ of a false alarm, we will determine the limitation level x_0 of the dimensionless $(U_{сш}/\sigma_{ш})$ amplitude $x(t)$ of the input signal:

$$P_{лт} = \int_{x_0}^{\infty} W_{ш}(x)\,dx = \int_{x_0}^{\infty} x \exp\left[-\frac{x^2}{2}\right] dx = \exp\left[-\frac{x_0^2}{2}\right].$$

Hence

$$x_0 = \sqrt{-2\ln P_{лт}}.$$

Quantity $a = \frac{U_c}{\sigma_{ш}}$ is determined by the assigned value for the probability $P_{по}$ of correct detection and by the calculated level x_0 of limitation:

$$P_{по} = \int_{x_0}^{\infty} x \exp\left[-\frac{x^2+a^2}{2}\right] I_0(ax)\,dx.$$

Quantity a, which ensures target detection with an assigned probability for the fixed probability of false alarms, serves as the threshold signal measure and is designated $a_{пор}$.

Then, in accordance with expression (3.2) l_0 will be defined as

$$l_0 = l(x_0) = \exp\left[-\frac{a_{пор}^2}{2}\right] I_0(a_{пор}x_0). \qquad (3.6)$$

3.3. Resolver for Separating a Packet of Pulse Signals

As a rule, circular scan radars operate in such a way that, during the scanning of an area, each target is illuminated repeatedly. Because of this the target signal is represented not as one reflected pulse, but a packet (packets) of pulses which have the same range but a different azimuth.

Target detection by packet gives the best results, since a considerably greater volume of information is used than in the case of a single signal. The repeated appearance of a pulse in the same scanning area substantially raises certainty in target detection.

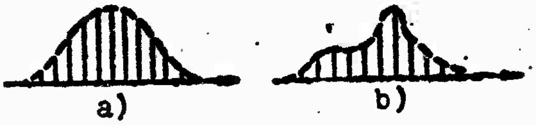


Fig. 3.3. Pulse packet envelope:
a) idealized; b) actual.

The envelope shape of the packet is determined by the type of function which describes the radar antenna radiation pattern (Fig. 3.3a). The actual envelope of the pulse packet (Fig. 3.3b) differs from the idealized. This distinction is explained by target fluctuation caused by the interference of signals reflected from its individual sectors during change in target orientation by natural and man-made interferences (noises), etc.

The coherent and incoherent reception of the packet pulses differ. With coherent reception there is a change in the phase from signal to signal, and this is taken into consideration during the processing of the packet. With incoherent reception there is no change in the phase from signal to signal. During processing the phase of every packet pulse is taken as a random quantity evenly distributed in the interval from 0 to 2π.

Subsequently, only the incoherent reception, with which the information contained only in the amplitudes of the pulse packet is used, is examined. Such information is the combined density of probability distribution of the pulse amplitudes in the packet.

If to process a single pulse signal it would have been possible to satisfy the probability densities of the $W_{\text{ш}}(x)$ one-dimensional noise samples and the "signal plus noise" mixture $W_{\text{cш}}(x)$, then to process the packet of pulse signals the corresponding probability densities of the n-dimensional samples are necessary:

$$W_{n\,\text{ш}}(x_1, x_2, \ldots, x_n/H_0);$$
$$W_{n\,\text{cш}}(x_1, x_2, \ldots, x_n/H_1),$$

which are determined at specific instants $t_1, t_2, \ldots, t_n$.

With the reception of the packet, since the time between samples is great, there is no correlation between them.

Consequently, the random magnitudes of the samples (pulse amplitude) are independent of each other, and the probability density of an n-dimensional sampling is equal to the product of the probability densities of the individual samples, i.e.

$$W_{n\,\text{cш}}(x_1, x_2, \ldots, x_n) = \prod_{i=1}^{n} W_{\text{cш}}(x_i);$$
$$W_{n\,\text{ш}}(x_1, x_2, \ldots, x_n) = \prod_{i=1}^{n} W_{\text{ш}}(x_i).$$

To obtain the detection algorithm of a radar packet (analogous with the case of a single pulse signal) let us make use of likelihood ratio expression (2.39):

$$l(x_1, x_2, \ldots, x_n) = \frac{W_{n\,\text{cш}}(x_1, x_2, \ldots, x_n)}{W_{n\,\text{ш}}(x_1, x_2, \ldots, x_n)} \geqslant l_0. \qquad (3.7)$$

The conditional probability distribution of the voltage envelope of the "signal plus noise" mixture and voltage envelope of pure noise for the individual pulses of the packet (individual samples) we assume to be known:

$$W_{\text{cш}}(x_i) = x_i \exp\left[-\frac{x_i^2 + a_i^2}{2}\right] I_0(a_i x_i) \qquad (3.8)$$

and

$$W_{m}(x_i) = x_i \exp\left[-\frac{x_i^2}{2}\right] \qquad (3.9)$$

In accordance with the formulas (3.8) and (3.9), expression (3.7) reduces to the following form:

$$l(x_1, x_2, \ldots, x_n) = \frac{\prod_{i=1}^{n} W_{cm}(x_i)}{\prod_{i=1}^{n} W_{m}(x_i)} = \prod_{i=1}^{n} \frac{x_i \exp\left[-\frac{x_i^2 + a_i^2}{2}\right] I_0(a_i x_i)}{x_i \exp\left[-\frac{x_i^2}{2}\right]} = e^{-\sum_{i=1}^{n} \frac{a_i^2}{2}} \prod_{i=1}^{n} I_0(a_i x_i) > l_0 \qquad (3.10)$$

After taking the logarithm and after transformations, expression (3.10) assumes the form of the algorithm for optimal detection of the pulse signal packet (Fig. 3.4):

$$\sum_{i=1}^{n} \ln I_0(a_i, x_i) > \ln l_0 + \sum_{i=1}^{n} \frac{a_i^2}{2}. \qquad (3.11)$$

This inequality also determines the structure of the resolver. From it it follows that the resolver should sum up the analyzed samples of the sampling values of the process after the functional transformation $z_i = \ln I_0(a_i x_i)$, and then the sum obtained is compared with the threshold which depends on the number of samples. With an increase in the total amplitude of the threshold level the decision on target detection is made.

A block-diagram of the resolver which realizes the algorithm in (3.11) is shown in Fig. 3.5. The use of analog accumulators makes it possible to perform automatic threshold detection of reflected signals without a substantial reduction in radar coverage.

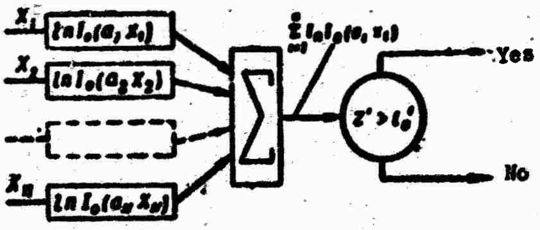


Fig. 3.4. Block-diagram of the detection algorithm of the pulse packet.

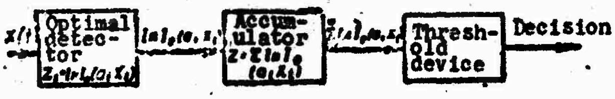


Fig. 3.5. Block-diagram of a resolver for detecting the pulse packet.

At least two types of equipment are possible for analog accumulation: charge-storage tubes and ultrasonic delay lines. These devices have serious disadvantages which hinder their use.

Thus, in charge-storage tubes to it, it is extremely complex to provide linearity in the accumulation law, which leads to losses in the threshold signal. The resolving power of charge-storage tubes thus far is still lower than the resolving power of radar, which leads to increased errors during coordinate evaluation. Difficulties arise also in connection with the necessity that with each scan the writing beam will pass over the same line on the surface of the accumulator plate. The readout beam should also pass along this line.

Substantial deficiencies are distinctive to ultrasonic delay lines. The basic of them are awkwardness (with large n), as well as the need for the precise coordination of delay with the repetition frequency of radar sounding pulses. For every new radar system the accumulator should be different. The use of accumulators on ultrasonic lines is generally impossible, if the repetition frequency of radar main bangs is variable.

The indicated deficiencies in analog accumulators induce the conversion to digital accumulators, similar to those which are used in digital computers (DC). Such accumulators are somewhat more simple, more economical and more reliable, and can be adapted for work with different radars by changing the accumulation program.

Digital detectors have one additional substantial advantage. They are monotype with the equipment by the computers which perform further processing of the radar information.

Meanwhile it is evident that the transition to digital methods of processing means a departure from those accepted in radar statistical processing methods. This causes the need for the digital transformation (quantization) of processable signals. This transformation unavoidably leads to losses in the threshold signal and therefore to losses in the quality of the processing.

Thus, the problem of the digital processing of radar signals in the unit which deals with automatic detection and evaluation of the target coordinates includes: 1) the transformation of the reflected radar signals to a form convenient for processing with the aid of digital computers; 2) the algorithmization of the processes in the processing of sampled radar signals; and 3) the realization of processing algorithms with the aid of digital computer technology.

3.4. Digitization of the Surveillance-Pulse Radar Signals

To process radar signals with the aid of digital equipment it is necessary to first convert the voltage, continuous with time and amplitude, of an useful signal at the output of the detector of the receiver into digital form.

The first stage in the digitization of the signals is time digitization. In this case a continuous (for observation time 0-$t_{н}$) signal is converted into a time sequence of digital amplitude values into sampling moments or into a digital sequence of other magnitudes connected with amplitude of the signal in the observation interval. If the signal amplitude is the function not of one, but several independent arguments (for example - time, Doppler frequency, etc.), then every argument should be digitized.

The second stage of digitization is the conversion of the amplitude sequence into a certain number of gradations, i.e., the representation of sampling amplitudes by a certain number in the decimal number system. Subsequently, the information on the amplitudes of the signal samples can be coded in a binary code.

The overall signal at the output of a radar receiver is a continuous function of time and target coordinates. Independent of the radar method and radar type, this signal during scanning develops into signal U(t), which depends only on time t. Simultaneously, natural digitization of the coordinates occurs at this argument during scanning.

As an example let us examine the scanning process of the signal of double-coordinate pulse radar with circular scan. The overall signal in this case is represented in the form

$$U_{\text{сш}} = U_{\text{сш}}(Д, β, t),$$

where Д, β - the two-dimensional coordinates of the target relative to the location of the radar station; $U_{\text{сш}}$ - the total voltage of the received signals and of the noises inherent in the receiver.

As a result of periodic scanning with period T_0, the continuous functions Д(t) and β(t) of the coordinates are converted into a digital series of instantaneous values Д(iT_0), β(iT_0) of the coordinates, where i is the scan number. This is the first stage of digitization with respect to time.

Finer digitization with time is obtained as a result of sending main bangs with period T_n. In this case the coordinate which coincides with the scanning plane (in our case coordinate β) is simultaneously digitized. Signals obtained during one period of sending main bangs, i.e., during one range sweep, remain continuous with time.

Thus, signal at the output of the receiver in pulse radar can be presented in the form of a time function:

$$U_{\text{см}}=U_{\text{см}}\left(iT_0,\ kT_n,\ \frac{ct}{2}\right),$$

where i - scan number; k - probing number within the limits of one scanning period; c/2 - range sweep speed.

Let us agree upon calling quantization both the digitization of signals in terms of amplitude, and time digitization. We will call the intervals of digitizing the signals with respect to time the time quantization intervals, and the intervals of digitizing with respect to amplitude - the amplitude quantization levels.

The functional circuit of radar signal quantization is given in Fig. 3.6, and the time diagrams which illustrate the work of the circuit - in Fig. 3.7. The voltage of the envelope for the "signal plus noise" mixture from the output of the radar receiver (curve 1) enters at one of the inputs to the time quantization circuit. The scale pulses of time quantization (with a duration of $\tau_{кв}$) with period t_k (curve 2) enter another input to the circuit. At the output of the time quantization circuit, equally discrete sample of the envelope voltage is obtained within the limits of each quantization pulse (curve 3). This sample enters at the input of the amplitude quantization circuit.

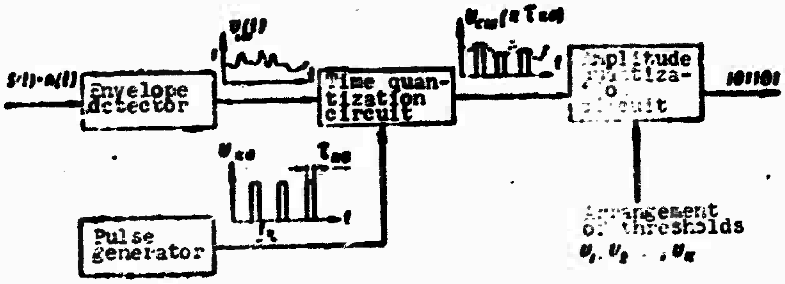


Fig. 3.6. Functional circuit of radar signal quantization

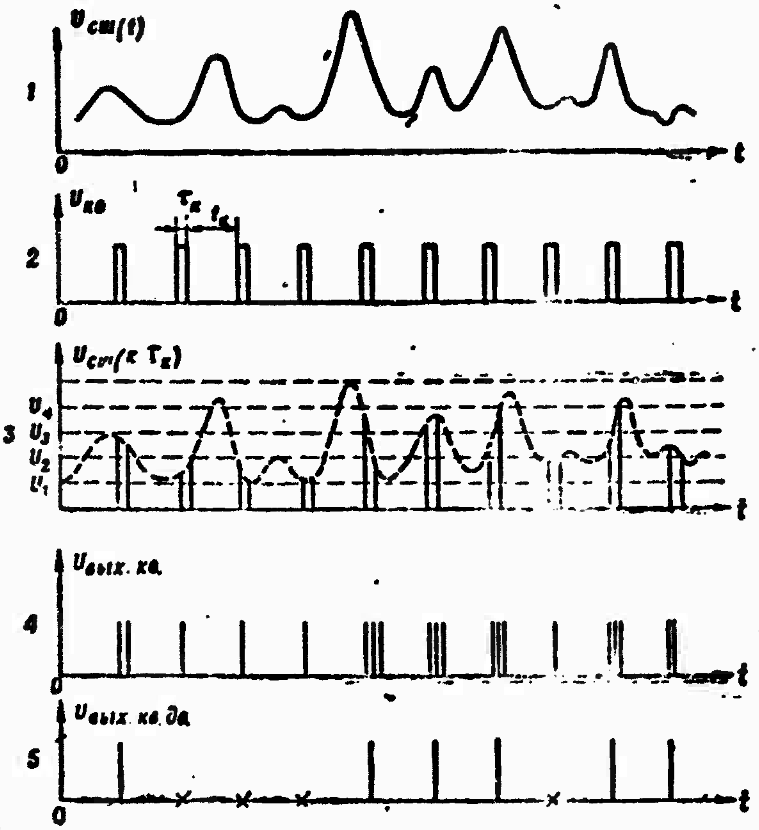


Fig. 3.7. Time diagrams which illustrate the process of radar pulse quantization.

The amplitude quantization circuit in general has m thresholds and m + 1 levels and issues upon output the level number, which is exceeded by the sample in each interval of time quantization. The level number can be coded in a binary code or represented by the standard amplitude pulse number, which is equal to the number of exceeded levels (curve 4). In the particular case of binary quantization (for example with threshold U_2), the circuit issues a pulse of standard amplitude and duration (unity), if the pulse amplitude exceeds threshold U_2, and a blank (zero), if this threshold is not exceeded (curve 5). Time quantization leads to a break in the range sweep of the radar at the elementary sectors with width

$$\Delta d = \frac{c t_{\text{к}}}{2},$$

where $t_{\text{к}}$ - time quantization interval.

The number of such sectors is determined from the expression

$$m = \frac{\text{Д}_{max}}{\Delta d},$$

where Д_{max} - maximum radar detection range.

In the process of the periodic sending of main bangs with the simultaneous rotation of the antenna, the radar visual range is split into elementary sectors by azimuth. The angular dimension of an elementary sector by azimuth (angular discreteness of the azimuth) is

$$\Delta\beta = 2\pi \frac{T_{\text{п}}}{T_0},$$

where $T_{\text{п}}$ - the main bang sending period; T_0 - the antenna rotation period.

$$N_s = \frac{2\pi R_{max}}{\Delta R \Delta\beta},$$

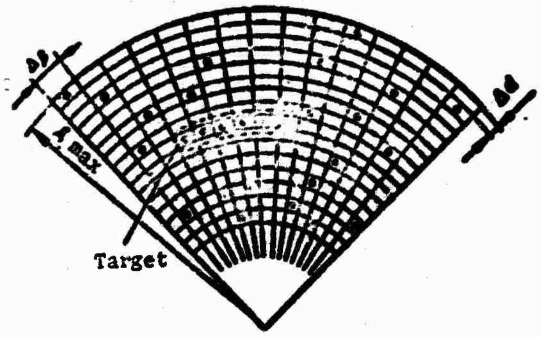


Fig. 3.8. Image of the radar visual range after the time and binary amplitude quantization of the signals.

The quantization of signal amplitudes leads to the digitization of the image "brightness" of each elementary cell. Depending on the output of the amplitude of the transmitted signal, this brightness will be characterized by the discrete number of quantization levels. During binary quantization a black-and-white image of the visual range is obtained, where the black background designates the presence of quantized signals (units) in the appropriate cells, and the white - their absence.

Processing time and amplitude of quantized signals reflected from a target in this case consists of the sequential analysis of the brightness of the cells belonging to one range (since pulses in the packet have the same range).

3.5. Binary Quantization of Signals

During the selection of the thresholds of the amplitude quantization of radar signals, used are the criteria connected with the loss of information on useful signal in the quantization process (information criteria), as well as the criteria connected with making a decision on the detection of the single signal or the packet of the signals, among which the basic are minimum risk criterion and the Neumann-Pearson criterion.

Let us examine the binary quantization of signals as the most simple. In this case the quantizer has only one threshold and the voltage on its output can take only two values: 0 and 1. Thus, signals at the output of the quantizer are the totality of random binary numbers (zeros and unities) which appear with probabilities P(0) and P(1), respectively. With a known distribution $W_{\text{сш}}(x)$ of single amplitudes and the set threshold x_0 of binary quantization, probabilities P(0) and P(1) are determined in the following manner:

$$P(1)=\int_{x_0}^{\infty} W_{\text{сш}}(x)\,dx;$$

$$P(0)=\int_{0}^{x_0} W_{\text{сш}}(x)\,dx.$$

To find the optimal threshold it is necessary to differentiate the expression for the average risk on threshold and to make the result zero. We will write the average risk for the detection of the single standardized signal in accordance with general formula (2.35) in the form

$$R=r_{\text{пц}}P_1\int_{0}^{x_0} W_{\text{сш}}(x)\,dx+r_{\text{лт}}P_0\int_{x_0}^{\infty} W_{\text{ш}}(x)\,dx. \qquad (3.12)$$

Let us take a simple case:

$$r_{\text{пц}}=r_{\text{лт}}=1,\ P_1=P_0=0{,}5.$$

Then, by differentiating expression (3.12) with respect to x_0, we will obtain the equation for determining optimal threshold $x_{0\,\text{opt}}$:

$$W_{\text{сш}}(x)\,|_{x_0=x_{0\,\text{opt}}}=W_{\text{ш}}(x)\,|_{x_0=x_{0\,\text{opt}}}.$$

In accordance with expression (3.13) optimal threshold $x_{0\,opt}$ should be chosen in the manner as shown in Fig. 3.9.

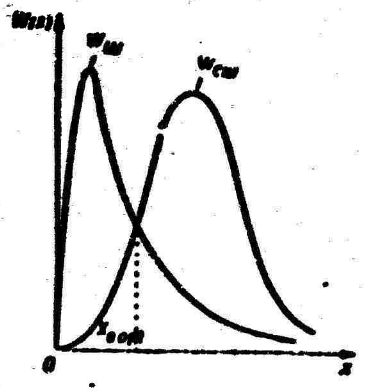


Fig. 3.9. For the selection of the optimal threshold of binary quantization by the average risk criterion.

Let us examine concrete examples.

Example 1. For the standardized amplitudes of the "signal plus noise" mixture at the output of a synchronous detector we have:

$$W_{сш}(x) = \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{(x-a)^2}{2}\right];$$
$$W_{ш}(x) = \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{x^2}{2}\right].$$

By substituting these expressions into equation (3.13), after simple transformations we will obtain

$$x_{0\,opt} = \frac{a}{2}.$$

Example 2. For the standardized amplitudes of a mixture of a non-fluctuating signal and interference at the output of an envelope detector we have:

$$W_{сш}(x) = x \exp\left[-\frac{x^2+a^2}{2}\right] I_0(ax);$$
$$W_{ш}(x) = x \exp\left[-\frac{x^2}{2}\right].$$

By substituting these expressions into equation (3.13), we get

$$\ln I_0(ax_{0\,opt}) = \frac{a^2}{2}. \qquad (3.14)$$

Equation (3.14) is not solved in clear form. In the case of weak signals ($ax_{0\ opt} \ll 1$), using the asymptotic expansion of the Bessel function:

$$I_0(ax_{0\ opt}) \approx 1 + \frac{a^2x_{0\ opt}^2}{4};$$

$$\ln I_0(ax_{0\ opt}) \approx \frac{a^2x_{0\ opt}^2}{4},$$

we will obtain

$$x_{0\ opt} = \sqrt{2}.$$

For the series of single amplitude values ($a \geq 1$) the numerical solution of equation (3.14) gives the following results (Table 3.1).

Table 3.1.

a	1	2	3	4
$x_{0\ opt}$	1.5	1.75	2.1	2.5

With increased single amplitude the optimal threshold approaches a:2.

Figure 3.10 gives the curves calculated from formula (3.12) of the average risk as a function of the threshold of quantization for certain ratios of signal to interference, which are most characteristic for circular-scan pulse radar. The curves have weakly expressed minimums. This confirms the noncriticality of selecting the threshold of binary quantization. On the average, for a sufficiently broad range of signal-to-interference ratios, the optimal thresholds lie within the 1.8-2.2 range, which makes it possible to choose a fixed threshold value for all expected signal-to-interference ratios without noticeable loss in the detection probability.

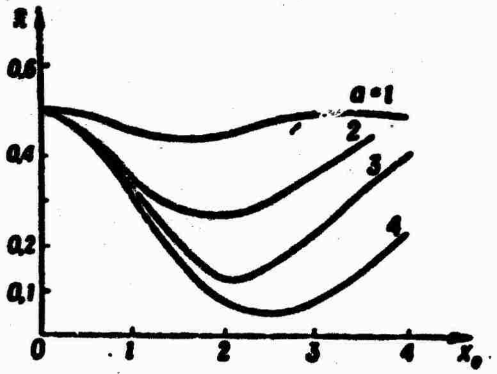


Fig. 3.10. **Average risk for the binary quantization of a single** signal.

3.6. Algorithm for the Optimal Detection of a Quantized Signal Packet

When one is examining the optimal processing of a quantized signal packet, it is necessary to bear in mind that quantization is unavoidably connected with loss of information. Therefore, theoretically no algorithm for the processing of quantized signals can ensure better (or even equal) qualitative characteristics, than the algorithm for the optimal processing of unquantized signals by analog computers. On the other hand, quantified signals can be processed with the aid of digital computers (digital accumulators) for which, unlike analog computers, the phenomenon of saturation is not characteristic. Therefore, in spite of losses during quantization, the discrete storage of quantized signals in the majority of cases will be no less effective than analog storage.

Let us examine the algorithm for optimal detection of the quantized signal packet with discrete storage. From the position of the statistical decision theory the problem of detecting a packet of binarily-quantized signals is formulated in the following form.

Let a discrete selection be made of the envelope values $\{x_1, x_2, \ldots, x_n\}$ on n neighboring elementary sectors by azimuth within the limits of one range ring (Fig. 3.8). Each of the sampling values of x_i is subjected then to binary amplitude quantization by means of comparison with threshold level x_0. The outcome of the unitary testing of x_i is considered positive, if the corresponding sampling value exceeds the threshold, and negative, if the threshold is not exceeded.

The total set of results after quantization is sequence of zeros and unities. This sequence enters the input of the resolver, whose task is to decide optimally on the basis of analysis of the obtained sequence of zeros and unities whether the made selection represents the packet of pulses reflected from the target or whether it pertains to the interference.

To solve the formulated problem the resolver should process the incoming signals in accordance with some preassigned algorithm. In this case the optimal algorithm of detection, as in the case of the unquantized signals, comes to checking hypothesis H_0^* about the absence of useful signal against alternative hypothesis H_1^* about its presence. To construct the detection algorithm it is first of all necessary to find the statistical characteristics of the zero and of unity sequence which is subject to processing.

Let us designate p_i as the probability of obtaining unity at the i-th position, and q_i - as the probability of obtaining zero at this position. Apparently, $q_i = 1 - p_i$. The probability of obtaining any of the two possible outcomes as a result of i-th testing can be written in the form

$$p(x_i) = p_i^{x_i} q_i^{1-x_i}, \quad (x_i = 0; 1).$$

Because of the independence of the tests the combined probability of obtaining some combination of the zeros and of the unities in all n tests equals

$$p(x_1, x_2, \ldots, x_n) = \prod_{i=1}^{n} p_i^{x_i} q_i^{1-x_i}. \qquad (3.15)$$

In accordance with formula (3.15) the probability of obtaining unities in each of the n tests is

$$p(\underbrace{1, 1, \ldots, 1}_{n}) = \prod_{i=1}^{n} p_i.$$

and the probability of obtaining all zeros is

$$p(\underbrace{0, 0, \ldots, 0}_{n}) = \prod_{i=1}^{n} q_i.$$

Now on the basis of formula (3.15) it is possible to easily write expressions for the likelihood functions of hypotheses H_1^* and H_0^*.

The likelihood function of hypothesis H_1^* takes the form

$$L(x_1, x_2, \ldots, x_n / H_1^*) = \prod_{i=1}^{n} p_{s_i}^{x_i} q_{s_i}^{1-x_i},$$

where p_{s_i} - the probability of obtaining unity at the i-th position of the signal packet ($q_{s_i} = 1 - p_{s_i}$).

The likelihood function of hypothesis H_0^* takes the form

$$L(x_1, x_2, \ldots, x_n / H_0^*) = \prod_{i=1}^{n} p_{N_i}^{x_i} q_{N_i}^{1-x_i},$$

where p_{N_i} - the probability of obtaining unity at the i-th position in the interference area; this probability is identical for all i.

Let us find the likelihood ratio and compare it with constant number l_0, which is selected in accordance with the established criterion of the optimality of the solution:

$$l(x_1, x_2, \ldots, x_n) = \frac{L(x_1, x_2, \ldots, x_n/H_1)}{L(x_1, x_2, \ldots, x_n/H_0)} = \prod_{i=1}^{n} \left(\frac{p_{S_i}}{p_{N_i}}\right)^{x_i} \left(\frac{q_{S_i}}{q_{N_i}}\right)^{1-x_i} > l_0$$

The logarithm of the likelihood ratio is

$$\ln l(x_1, x_2, \ldots, x_n) = \sum_{i=1}^{n} \left[x_i \ln \frac{p_{S_i}}{p_{N_i}} + (1 - x_i) \ln \frac{q_{S_i}}{q_{N_i}} \right] > \ln l_0 \qquad (3.16)$$

After simple transformations of expression (3.16) we will obtain

$$\sum_{i=1}^{n} x_i \ln \frac{p_{S_i} q_{N_i}}{p_{N_i} q_{S_i}} > \ln l_0 - \sum_{i=1}^{n} \ln \frac{q_{S_i}}{q_{N_i}}. \qquad (3.17)$$

By designating in expression (3.17)

$$\eta_i = \ln \frac{p_{S_i} q_{N_i}}{p_{N_i} q_{S_i}}, \qquad \ln l_0 - \sum_{i=1}^{n} \ln \frac{q_{S_i}}{q_{N_i}} = \text{const} = z,$$

we will write the final formula for the algorithm of the optimal detection of the quantized signal packet:

$$\sum_{i=1}^{n} x_i \eta_i > z. \qquad (3.18)$$

Function η_i, which takes into consideration the expected values of the probability of obtaining unities and zeros in the signal and interference area, is called the weight function of binarily-quantized signal detection. The envelope of the function corresponds in form to the envelope of the radiation pattern of the radar antenna.

In accordance with algorithm (3.18) the process of optimally detecting the binarily-quantized signals packet comes to the fulfillment of the following operations: to the storage of signals (unities and zeros) at the output of the quantizer within the limits of packet width (at n positions); to adding up the values of the programmed weight function at those positions at which the unities from the output of the quantizer are obtained, and to the comparison of the obtained sum with threshold number z and to the issue of a decision on the detection (or nondetection) of the packet.

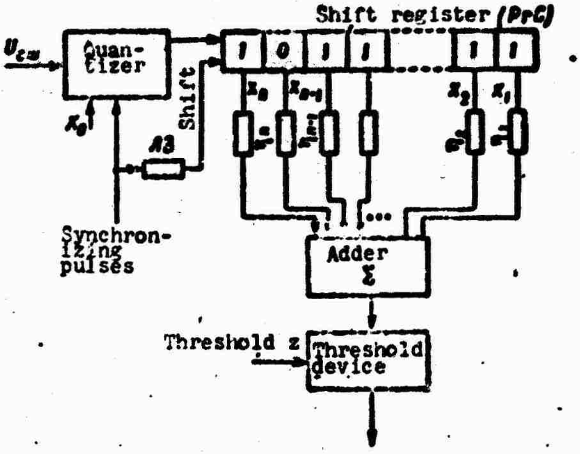


Fig. 3.11. Structural diagram of the resolver for the detection of the binarily-quantized signal packet.

The structural diagram of the detection resolver, which realizes the indicated algorithm, is illustrated in Fig. 3.11. For simplicity, in the diagram the memory of the quantized signals is presented in the form of a shift register (SR). The weight function is conditionally programmed in impedances. The accumulator is the adder of currents or voltages. The enumerated elements of the circuit can be realized by the means of digital computer technology.

The characteristic feature of the resolver for the detection of quantized signals is the presence of two thresholds: the first is determined in the quantizer and is the input threshold (the detection threshold of the individual pulses); the second is determined at the output of the adder circuit and is the detection threshold of the packet. Let us assume that probabilities p_{s_i} are identical within the limits of the entire width of the radiation pattern (the packet has a rectangular shape). Then, from expression (3.17) it is possible to obtain the optimal algorithm for the detection of the rectangular-shaped packet:

$$\sum_{i=1}^{n} x_i > \frac{\ln l_0 - n \ln \frac{q_S}{q_N}}{\ln \frac{p_S q_N}{p_N q_S}}, \qquad (3.19)$$

where right side - a constant value.

As can be seen from expression (3.19), in the case of the rectangular packet the procedure for detecting the quantized signals comes to the pulse counting of the unities obtained at the output of the quantizer within the limits of packet width and to comparison of the number of accumulated pulses with a certain threshold number. The detection circuit is simple and realizable with the aid of a conventional binary counter.

For simplicity in detection resolvers, binary counters are often used also in the processing of real (nonrectangular) packets. However, such resolvers will no longer be optimal, since during processing the real shape of the radiation pattern of the radar antenna is not taken into consideration.

3.7. Principle Weightless Processing of Radar Signals

The examined methods of processing packets of quantized radar signals are still rather complex in realization. The basic complication in the circuits is connected with the need for programming weight functions for packet detection and determination of their azimuth position.

For the purpose of simplification the equipment packet is considered to be rectangular. In this case the sole characteristic by which it is possible to distinguish the target area from the interference area is the increased density of unities within a certain interval of positions (for example, within the limits of the radiation pattern width of the radar antenna). The density of the unities in the target area will naturally be greater than the density of the unities in the interference area. Consequently, weightless processing equipment should estimate the density of the unities and react to density changes. During processing, in terms of the density of the unities, the basic problem is the fixation of the packet boundaries. To do this certain rules, by which the affiliation of the individual quantized signals to the same packet is determined, are established in advance.

The rules of fixation have four criteria: the criterion of the fixation of the beginning of the packet, the criterion of the fixation of the end of the packet, the criterion of detection, and the criterion of the fixation of the middle of the packet.

The totality of the criteria assigned in descriptive form or in the form of corresponding mathematical and logical expressions determines the algorithm for the weightless processing of the packet.

It is possible to take the appearance of a certain total set of unities and zeros at a fixed interval of positions or a certain series of the unities as the criterion of the fixation of the packet beginning. Furthermore, it can be required that the beginning of the packet be fixed immediately after obtaining the first unity.

For example, the presence of a single unity pass at the next position can be taken as the criterion of the fixation of the end packet. In this case the packet represents the totality of the positions at which the unities follow without passes. The end of the packet can also be fixed in terms of the presence of a series of two or several (k) passes in succession. Then the packet is the totality of the positions at which k - 1 and less passes are possible. It is possible finally to make arrangements to fix the end of the packet in terms of the appearance of some combination of zeros and unities at an assigned number of positions.

Most often the end of the packet is fixed in terms of the presence of a series of k passes in succession. When the number of passes is being selected, it is necessary to keep the following in mind. If this number is small, then the probability of the "splitting" or even the loss of the target is great. With large k the azimuth resolving power and the accuracy of its measurement are impaired. For the correct selection of number k it is necessary to have experimental data on the number and distribution of passes in the packets during the tracking of real targets.

The accumulation (calculation) of unities between the beginning and the end of the packet is used for the detection of packet. The accumulator can be made so that the number of passes k - 1 in succession will also be less restored at the internal positions. In this case, in the detection process not the units, but the positions between the beginning and the end of the packet accumulate at the output of the quantizer. If the number of positions in the packet exceeds a certain threshold number z, a decision on the detection of the packet is made. Conventional binary counters can be used as accumulators. This simplifies the detection circuit.

The principle of pulse accumulation within the boundaries of the packet is illustrated by time diagrams in Fig. 3.12. In the absence of pulses at two adjacent positions, the accumulation of pulses ceases. The detection signal above the dotted line) is transmitted upon the accumulation of five pulses.

Besides signals on the position of the packet boundaries and on its detection, the accumulator can give out the number of pulses accumulated at the moment of fixation of the packet end.

If the beginning of the packet is fixed in terms of the presence of a certain series of m positions or in terms of some combination of l unities at m adjacent positions, the criterion for the beginning of the packet is also synchroneous with its detection criterion, and the end of the packet is fixed only for the subsequent determination of target azimuth. Such detectors are called programmed

$$„m/m - k" \text{ or } „l/m - k",$$

where the fraction denotes the detection criterion, and k - the criterion of the fixation of the packet end.

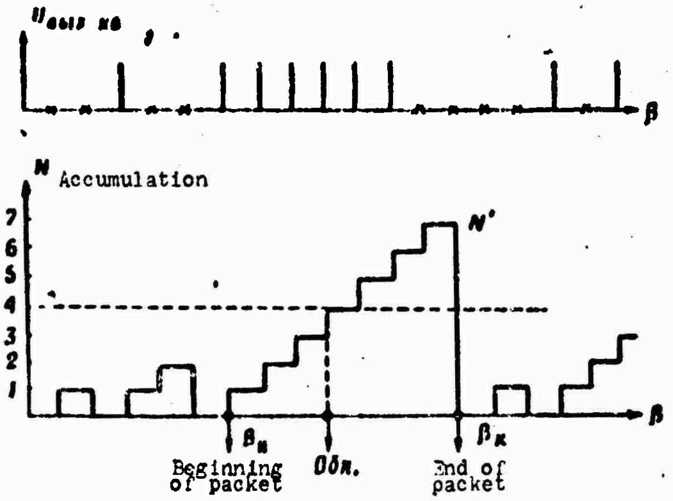


Fig. 3.12. Time diagrams which illustrate the principle of pulse accumulation between the adjacent series of two zeros.

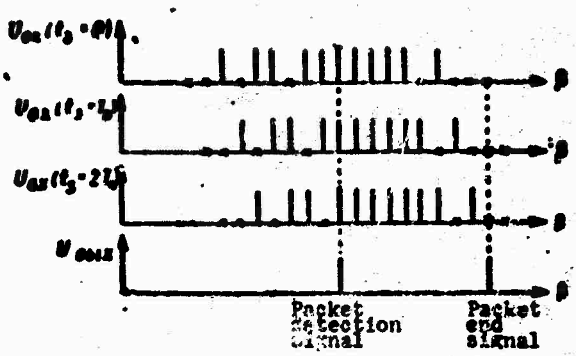


Fig. 3.13. Time diagrams which illustrate the principle of the fixation of the packet boundaries by programmed detector "3/3-3."

The principle of the fixation of the packet boundaries by programmed detector "3/3-3" is illustrated in Fig. 3.13. In this case the packet detection signal is transmitted upon the appearance in the input sequence of the quantized signals of three unities in succession (a series of three unities). The end of the packet is issued upon the appearance of a series of three zeros.

3.8. Methods of Estimating Target Azimuth During Weightless Processing

During weightless processing of radar signals the azimuth of the target is determined by means of fixing the position of the middle of the detected packet. The detector signals are used as signals by which the position of the middle of the packet is determined. In accumulators (Fig. 3.12) such signals are the pulses transmitted by the circuit during the fixation of the beginning and the end of the packet, and the number of positions between the beginning and the end of the packet (the width of the packet).

The position of the middle of the packet can be determined by the following methods.

1. By the azimuths of the pulses which fix the beginning and the end of the packet, with the subsequent determination of the arithmetic mean. The algorithm for estimating the azimuth in this case takes the form

$$\beta_0^* = \frac{\beta_{\text{н}} + \beta_{\text{к}}}{2}, \qquad (3.20)$$

where β_0^* is the target azimuth estimation, β_H is the azimuth of the beginning pulse of the packet, and β_H is the azimuth of the final pulse of the packet.

2. By the azimuth of the beginning and final pulse and the width of the packet. The algorithm for estimating the azimuth in this case takes the form

or

$$\left.\begin{aligned}\beta_0^* &= \beta_{\text{к}} - \frac{\Delta\beta N}{2}\\ \beta_0^* &= \beta_{\text{н}} + \frac{\Delta\beta N}{2},\end{aligned}\right\} \qquad (3.21)$$

where N is the number of positions from total number of positions n, which corresponds to the width of the detected packet, and $\Delta\beta$ is the angular discreteness of pulses in the packet.

Algorithms (3.20) and (3.21) are extremely simple in realization. In this case their advantage is in comparison with the algorithm for estimating azimuth with the aid of the weight function.

The weightless processing of radar signals employs digital computers for which the input signals are the sequences of the quantized radar signals which take only two values (zero and unity). These devices function according to a definite, predetermined algorithm and give out discrete signals on the detection of the packet and its boundaries.

Chapter 4

AUTOMATIC EXTRACTION AND CODING OF TARGET COORDINATES

4.1. Conjugation of Radar with Computers

During the initial period of radar development the basic method of determining coordinates issued by a detection station consisted of the following.

The operator determined the distance to the target and the azimuth and relayed these data by telephone. Later to increase reading accuracy, electronic range and azimuth markers began to come into use. Also, sector indicators, which give target blips in larger scale and by the fact increase the accuracy of determining coordinates, began to be used.

However, such a method had a number of deficiencies: poor accuracy in determining coordinates, which largely depends on the quality and the ability of the operators; every indicator indicated targets found in small volume, and that is why the coordinates of only one or a small number of closely situated targets were determined; the rate of data transmission is low, during transmission the data were substantially delayed and lost their value. These deficiencies, on one hand, and the great capabilities of radar in the rapid detection of a large number of targets, on the other, caused the development of automatic extraction equipment. These devices can be made in the form of analog servo systems or in the form of digital (discrete) extraction equipment.

The automatic extraction of data and the automatic input of them into a computer system solve the problem of the complete automation of those control circuits which have radar. Man retains only the function of checking and making the final decision on the basis of the machine-prepared data.

Analysis of the situation in the visual range of one or several radars with consideration for the cluster spacing and the simultaneous movement of many targets, is carried out only by a high-speed computer.

With the joint operation of the radar and the computer, the latter performs functions analogous to the functions of the operator. In this case the terminating device of the radar is not the indicator, but the automatic data extraction equipment.

Figure 4.1 illustrates the functional diagram of a radar station which outputs target coordinates as a binary code. The station includes: a synchronizer, a transmitter, an antenna switch, a receiver and scopes, and furthermore, automatic data extraction equipment, consisting of a preselector and devices for the digital determination of range Д and azimuth β.

The automatic data extraction device performs two basic operations.

The first operation is the isolation of a useful signal against the background of disturbances (noises), processing the signal in terms of threshold and in terms of position. This operation pertains to the primary processing of the signal and is performed by the preliminary selection unit or in brief form, the preselector.

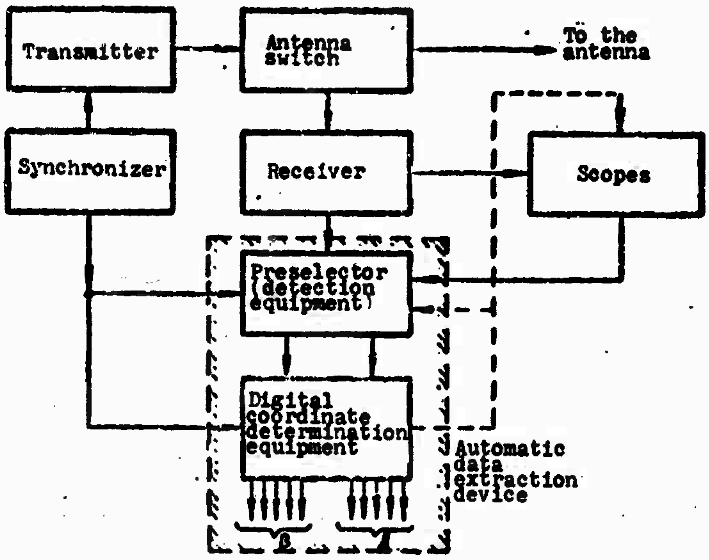

Transmitter
Antenna switch
To the antenna
Synchronizer
Receiver
Scopes
Preselector (detection equipment)
Digital coordinate determination equipment
Automatic data extraction device
β
Д

Fig. 4.1. Functional diagram of a radar station which outputs target coordinates in digital code.

The interferences are finally suppressed in the computer itself. The preselector performs only the preliminary extraction of the signal, from which it received its name.

The second operation is performed by equipment for determining the range coordinates and azimuth of the target. The range is determined by measuring the time interval from the moment the transmitter starts to the moment the reflected signal arrives, with the simultaneous conversion of the size of the interval into a binary code.

Azimuth is measured by means of determining the position angle of the antenna in binary code at the moment the packet of reflected signals arrives from the target.

All the found target coordinates enter the storage unit of the computer in the form of binary code numbers. This operation is similar to the operation of target coordinate reading by the operator according to the scale of the scope screen.

The automatic data extraction device, unlike the scope, does not reproduce situation for the entire scanning cycle. This function is performed by the storage unit of the computer itself.

Thus, the functions of isolating the signal, measuring coordinates and memorizing the situation during the work of the radar are separated in the computer while during the visual extraction of data all these operations are performed by the operator.

Because of such separation and computer use, each of the functions is individually performed more qualitatively and more precisely.

4.2. Digital Device for Detecting Radar Signals (Preselector)

Preselector (Fig. 4.2), using the periodic properties of the pulse packet, isolates the useful target signals against the background of interferences.

The pulses are fed from the output of the receiver to the limiter, where the voltages of the signals are restricted at the level of the established threshold (the binary quantization of the signals is performed). The signals which exceed the threshold of limitation pass through the limiter and are standardized to constant amplitude and pulse width with the help of the shaping cascade, which outputs a sequence of quantized pulses.

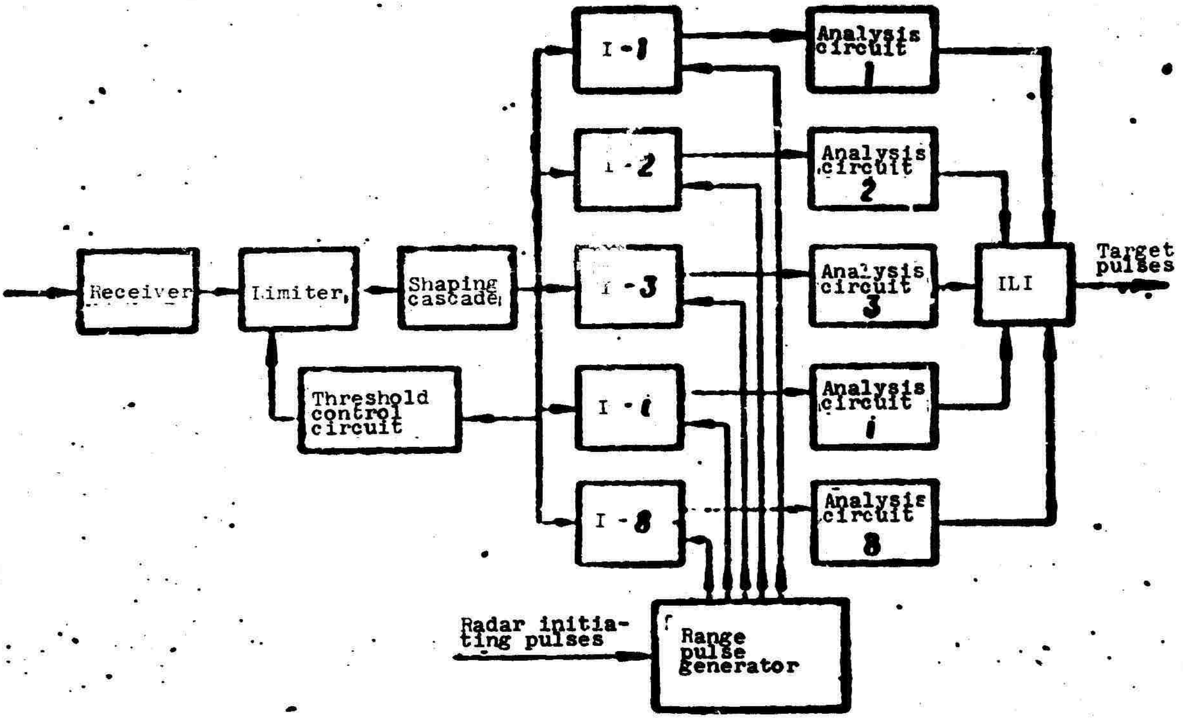


Fig. 4.2. Functional diagram of the digital detection device (preselector).

The limitation level is determined by the permissible false alarms. Since the absolute value of noises at the receiver output can vary (due to change in power supply modes, tube parameters and other reasons), then to maintain a constant relative level of limitation a circuit for controlling the threshold of limitation is used.

The standardized pulses enter the threshold control circuit input, and from the output there is taken constant voltage whose magnitude is proportional to the average number of pulses per unit of time. The standardized pulses also enter a series of logic elements: I-1, I-2, etc. Each logic element is a coincidence selector which gives a pulse at the output, if the pulses are fed simultaneously to its two inputs. Its range pulse is fed to one of the inputs of each coincidence selector. The range pulses have an identical pulse width and each subsequent pulse is displaced relative to the previous one for a time equal to the pulse width of one range pulse. The sequence of range pulses covers the entire range of radar coverage.

Each logic element is exposed in the same range interval and therefore pulses only in the assigned range interval can be at the output of each one.

Pulses from the output of the logic elements are analyzed by the pulse-sequence analysis circuit in accordance with the required law governing (logic) processing.

Pulses from the output of the analysis circuits, which follow at different instants, are fed to one logic element ILT, which gives a pulse at the output each time a pulse appears at any one of its inputs. The target pulses which are cleared of the noises and are normalized in terms of amplitude are taken from the output of this element.

The range pulse generator (Fig. 4.3) has a clock-pulse generator, a pulse counter, and a diode matrix. The circuit begins to operate under the action of radar synchronizing pulses which initiate the expander. The latter generates pulses whose pulse width corresponds to the time it takes for the electromagnetic energy to reach the maximum coverage and return (curve 2).

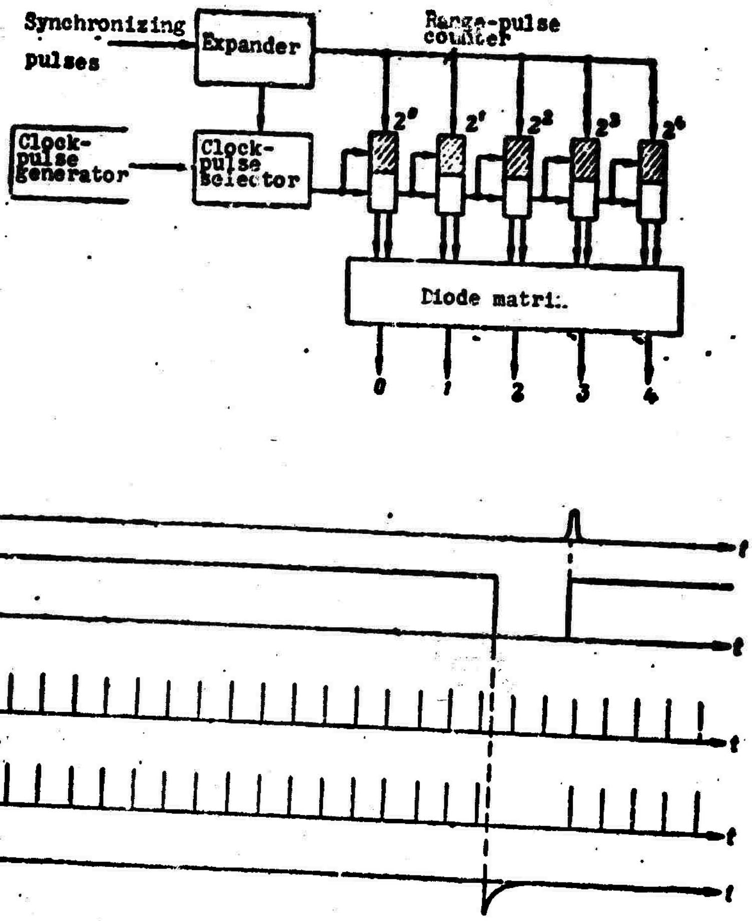


Fig. 4.3. Functional diagram of the range-pulse generator and curves of the voltages which control the range counter.

The clock-pulse generator is the source of standard pulses (curve 3). The pulse-repetition period determines the width of each range ring.

An expanded pulse (the strobe) turns on the calculation pulse selector, through which a series of pulses (curve 4) enters the pulse counter.

Every cell (flip-flop) of the pulse counter represents a discharge of a binary number, therefore the incoming series of pulses is fixed by the counter in binary numeration.

At the termination of the strobe a short pulse is produced (curve 5). This pulse recovers all the cascades of the pulse counter to the zero position ("reset"), preparing the counter for a new cycle of pulse series calculation in the following sending period.

Each discharge of the pulse counter is directly connected with the diode matrix, which transforms the numbers being issued by the counter into a sequence of range pulses. The circuit of the diode matrix which outputs eight sequential pulses is presented in Fig. 4.4. It contains as many pairs of horizontal lines, as the range counter has discharges (the figure shows three pairs of horizontal lines). Each pair of lines is designated "0" and "1" which correspond to two possible states of each digit of the counter. The high potential at the output of the flip-flop (the line is not connected to the ground) corresponds to unit (1), the low (the line is grounded) - to zero (0).

Prior to the admission of the first pulse into the counter input, all the discharges of the counter have the state 0, switches P (Fig. 4.4) of the diode matrix are in the upper position. Constant voltage U is fed through resistors R_0 and R_1 to the output of the circuit designated "0." Upon the arrival of the first range pulse, the counter of the first discharge is converted into the state "1," in this case voltage U will be applied through resistor R_2 to the first output, etc.

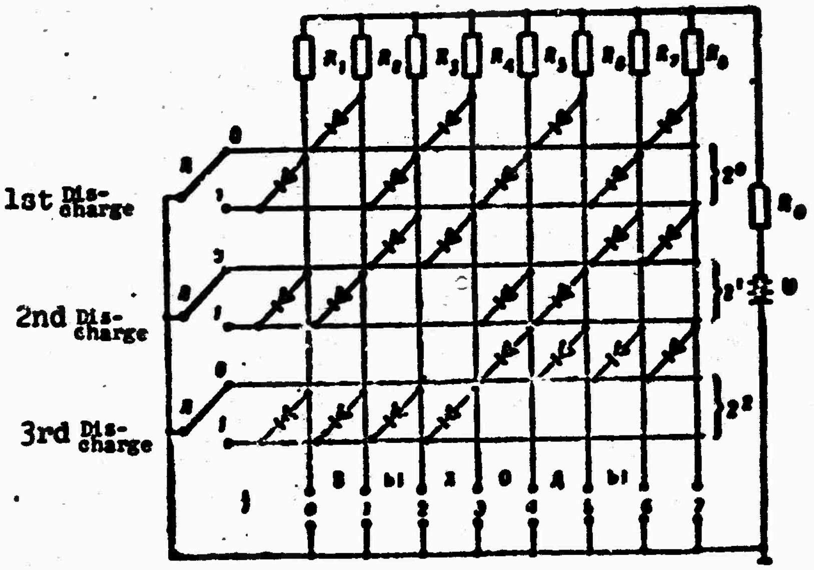


Fig. 4.4. Diode matrix of the range pulse generator. [П = P]

With the arrival of, for example, five pulses the state of counter can be characterized by number 101, which corresponds to position "1," of the switch for the third digit, to the zero position of the switch for the second digit and to position "1" of the switch for the first digit. Voltage U in this case will be fed through the resistor R_6 to the fifth output of the matrix.

Thus, at the outputs of diode matrix 0-7 we will obtain the sequential pulse series which corresponds to the states of the counter (Fig. 4.5). These pulses are fed to logic circuits I-1-I-8 (Fig. 4.2).

More complex is the device for analyzing the pulse sequence in the range ring (Fig. 4.6), it analyzes the incoming sequences of pulses in accordance with the assigned processing logic.

The circuit works by the following logic: one quantized pulse - the beginning of the packet, the number of passes in the packet - no more than two in succession; the detection pulse is issued by the circuit, if these conditions are satisfied for no less than eight pulse-repetition periods; the end of the packet - three passes in succession.

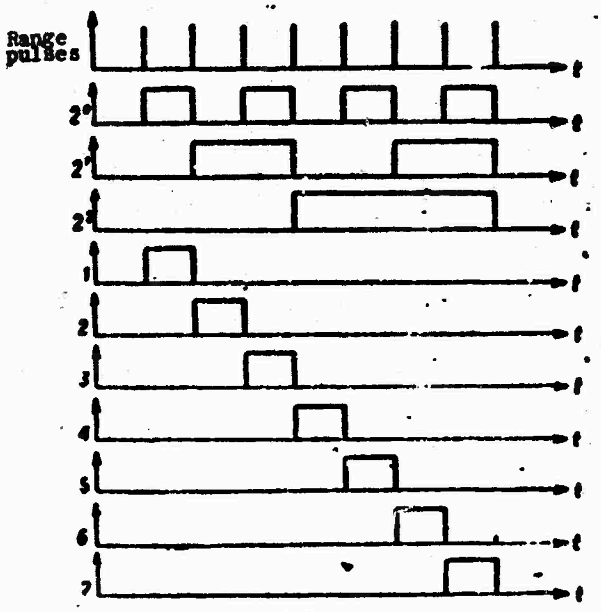


Fig. 4.5. Basic voltages of the range-pulse counter and diode matrix.

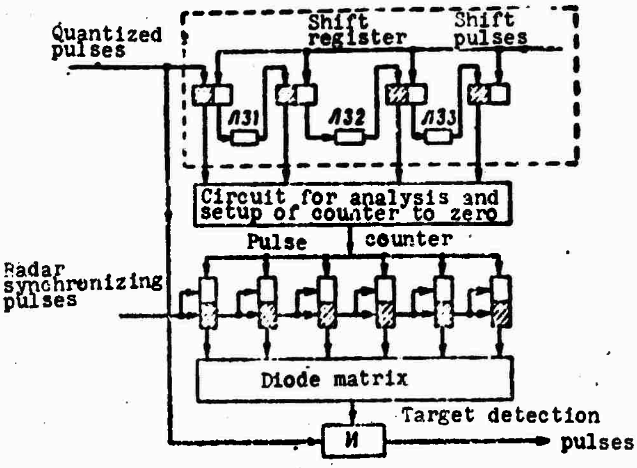


Fig. 4.6. Analyzer of the pulse sequence in the range ring. [И = I; ЛЗ = LZ].

A sequence of normalized pulses, caused by the presence of target signals in noises or only by noises is fed to the input of the circuit. The circuit should analyze the incoming sequence of pulses and issue a target detection pulse, if there is a target, and should not issue a detection signal, if the pulse sequence is formed by the noises.

The analyzer (Fig. 4.6) has a shift register, a circuit for analysis and for setting the counter to zero, a pulse counter and a diode matrix.

The shift register memorizes (registers) the four outcomes of checking the presence of quantized pulses in the repetition periods which follow each other. A distinction of the shift register from an ordinary pulse counter is only that the output of one cell (flip-flop) is connected to the input of the next cell, not directly, but through the delay line.

Shifting pulses perform the role of the zero-setup pulses; they put all register cells into the zero cells. If any cell of the shift register were in the zero state, then with the effect on it of the shifting pulse it would remain in this state. If this cell were in state "1," then after the admission of the shifting pulse it would be converted into state "0." In this case a pulse is produced in the cell; in a time equal to the transit time in the delay line, this pulse approaches the next cell, converting it into state "1."

Consequently, after the admission of the shift pulse state "1" will displace from the previous cell to the next.

For example, if in the shift register there is the binary number 1111, then to shift it to one digit to the right, it is necessary to feed one shift pulse to the shift line. Since this pulse is the zero-setup pulse for all the cells, they all simultaneously return to state "0." In this case at the outputs of all the cells there appears a positive pulse of voltage which, running through the delay line, enters the input of the subsequent cell and transfers it to state "1."

Thus, after the passage of one shift pulse instead of the code of number 1111,the code of number 0111, etc., will be fixed in the register.

If quantized pulses are not fed to the input of the shift register for a long time, state "0" is established in all cells of the shifting register.

With the appearanceof quantized pulse in any radar repetition period, the first cell converts to state "1," and when the shifting pulse is incoming, state "1" is displaced into the second cell. The shift pulse enters at the end of the repetition period; therefore the shift will already occur toward the next repetition period and the first cell will be prepared to receive the next pulse.

The results of registering the pulses in four sequential repetition periods enter the analyzing circuit from the shift register. This circuit analyzes the state of the cells in the shift register and in accordance with their state issues or does not issue the pulse to set the counter on zero.

With the four cells of the shift register it is possible to consider the state of the signals in four sequential repetition periods or less. Let us suppose the signals in cells 2, 3, and 4 will be subject to analysis, and cell 1 is used for intermediate memorization. In accordance with logic in processing the signals, the analyzing circuit will work in the following manner: if in cells 2, 3, and 4 the state is "0," the analyzing circuit gives the pulse for setting the counter on zero; if only one cell has a state of "1," the pulse counter during this period registers unity. The pulse counter computes the radar trigger pulses, and if the zero adjustment pulses do not enter, it continues to count. Let us illustrate the operation of the circuit by two examples.

One quantized pulse caused by noise overshoot is fed to the input of the shift register. In this case the state of the circuit elements can be presented in the form of Table 4.1.

Table 4.1.

Repetion period number	1	2	3	4	5	6
Presence of quantized signal	0	1	0	0	0	0
Counter reading	0	1	2	3	0	0

In the second repetition period the quantized pulse converts the first cell of the shift register to state "1," the shift pulse moves state "1" into cell 2, and cells 3 and 4 have state "0" as before. The pulse for setting the counter on zero will pass through the analyzing circuit, and counter will fix the first unit.

In the third repetion period the state of cells 2, 3 and 4 of the shift register will be characterized by code 010. The analyzing circuit will not give out a pulse to set the counter on zero, and counter will fix digit 2.

In the fourth repetition period digit "3" will be fixed, while in the fifth the counter will be set on zero.

In a more complex case (Table 4.2) the circuit begins the count with the first quantized pulse, counting continues in the presence of one or two zeros in succession, and the count ends when three zeros appear in succession. If before setting the counter on zero we fix its readings and the values of the end of the packet, it is possible to determine the middle of the pulse packet and, consequently, the azimuth of the target also.

Table 4.2.

Repetition period number	1	2	3	4	5	6	7	8	9	10	11	12	13	14
Presence of quantized signal	0	1	0	1	0	0	1	1	0	0	1	0	0	0
Counter reading	0	1	2	3	4	5	6	7	8	9	10	11	12	0

However, in order to avoid the erroneous measurement of target azimuth with solitary overshoots caused by noise, one uses a device which issues target detection pulses only in a case where after the arrival of the first quantized pulse, the condition about the number of passes (no more than two) is satisfied for a sequence of 8, 9, 10 or more repetition periods.

Such a device is made in the form of a diode matrix switched on to the first four digits of the counter (Fig. 4.7).

If for seven repetition periods the requirement for the sequence of quantized pulses is satisfied (the number of passes is no more than two), then at the beginning of the eighth period the pulse counter will fix eight units in binary numeration (1000). This corresponds to position "1" of the switch for the fourth digit of the diode matrix and to position "0" of the switches for the first three discharges.

In this case voltage U will be fed to circuit I (Fig. 4.6) through resistors R_0 and R_8 and the output of the 8-th matrix.

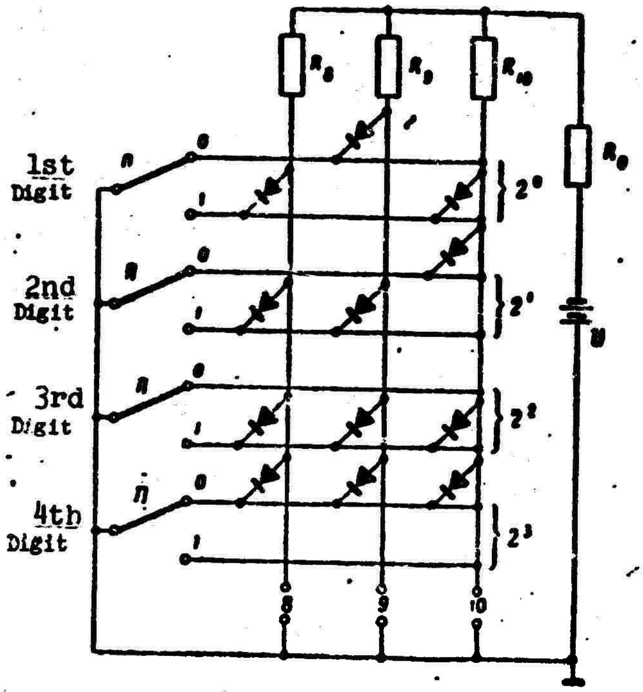


Fig. 4.7. Diode matrix which gives output voltages for 8, 9, and 10 clock pulses. [П = P].

In circuit I it will be retained for the entire radar repetition period, if the pulse for setting the counter on zero does not follow. With arrival of the quantized pulse in the eighth period at circuit I, the pulse proceeds to the output and will be the detection pulse delayed for a time corresponding to the time lag of the signal from target.

The output voltage at points "9" and "10" is used in a case, where it is necessary to use the 9th or 10th pulse for the selection of the quantized signal.

Such a construction of the circuit in at least eight repetition periods satisfies the condition that the number of passes does not exceed two, and the number of pulses in a packet for ten repetition periods of radar trigger pulses comprises no less than four.

The pulse sequence analyzing circuit (Fig. 4.6) is intended for operation in only one range ring. From the analyzing circuits which operate in all the range rings, the target detection pulses enter logic circuit ILI (Fig. 4.2) and then the ranging circuit - a digital ranging system.

4.3. Ranging in Digital Code

The basis of constructing a digital ranging system intended for output of target range in the form of a binary code is the dependence of range on the time lag of the signal reflected from the target:

$$\text{Д} = \frac{c\tau}{2},$$

where τ - the time lag; $c = 3 \cdot 10^8$ m/s.

This dependence is used to present range in the form of a binary number equal to the number N of pulses incoming from the clock-pulse generator for time τ:

$$N = F\tau,$$

where F - the pulse repetition frequency.

Then range

$$\text{Д} = \frac{c}{2F} N. \qquad (4.1)$$

The functional circuit of a digital ranging system which operates by this principle is depicted in Fig. 4.8. The circuit has a clock-pulse generator, a pulse counter, an expander (strobe generator), clock-pulse selector I-1 and distributor-shifter. All digits of the counter at the beginning of radar triggering are found in state "0."

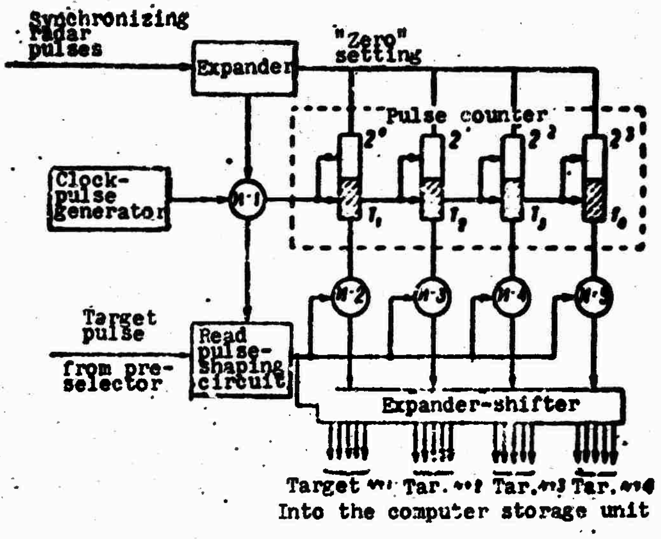


Fig. 4.8. Functional circuit of the digital ranging system. [И = I].

With the arrival of a synchronizing pulse which fixes moment of main bang radiation by the radar transmitter, the expander produces pulse (strobe), whose length corresponds to the maximum range of radar coverage. This pulse opens clock-pulse selector I-1 and the series of ranging pulses enters the counter. Counter computes the number of pulses having passed, and therefore target range. However, thus far these data from the counter do not enter anywhere, since the selectors of the range-code read-out - logic elements I-2, I-3, I-4 and I-5 - are included at the output of the flip-flop of every digit.

Only with the arrival of the target pulses from the preselector (through the read pulse-shaping circuit) are the code reading selectors opened for a time equal to the width of the target detection pulse and the computed number is output to the computer storage unit.

The read pulse-shaping circuit plays a special role in this system. This is caused by the existence of moments of time, when the range-code reading of the pulse counter is undesirable.

Such moments of time come right after feeding the next clock pulse (curve 1, Fig. 4.9), since after its arrival in the flip-flops of the counter, the transient processes which hinder the reading appear.

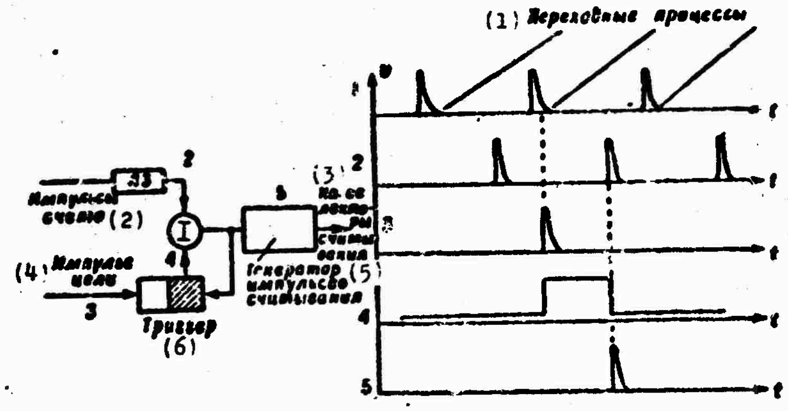


Fig. 4.9. Read pulse-shaping circuit and the stress curves which illustrate its operation.
KEY: (1) Transient processes; (2) Clock pulses; (3) To the read selectors; (4) Target pulse; (5) Read pulse generator; (6) Flip-flop.

Actually, let us assume that the reading occurs at the moment number 0011111 (31) changes to number 0100000 (32), i.e., at the moment when the units are changed by the zero, and the zero by the unit. Error can be allowed in every digit as a result of the reading. The minimum readable number - zero (zeros are considered in all digits), the maximum - 63 (the units are considered in all six digits). Thus, instead of number "31" or "32" an arbitrary number within the limits from 0 to 63 is issued!

The read pulse-shaping circuit provides range-code reading after the termination of the transient processes, even if the target detection pulse arrives at a point in time, when transient processes are not yet finished (curve 3, Fig. 4.9).

For this the clock pulses which pass delay line L3 and further enter coincidence circuit I are fed to the input of the circuit. However, in the initial state the coincidence circuit is closed at the second input by the low output potential of the flip-flop and the read pulse generator therefore does not operate. With the appearance of a target detection pulse the flip-flop is tripped supplying positive voltage to the coincidence circuit. The next clock pulse passes through the coincidence circuit and starts the read pulse generator.

A pulse (curve 5, Fig. 4.9) shifted relative to the appropriate clock pulse is formed at the output of the circuit. A pulse is fed to the read selectors. The time of delay line L3 is greater than the time of the transient process in the counter. This ensures correct reading of the range code. The flip-flop is returned to its initial state by the pulse from the output of the coincidence circuit, which is fed to the second input of the flip-flop.

With the termination of the pulse (the strobe) produced by the expander (Fig. 4.8), a short pulse which returns all the cells of the counter to the zero position is formed.

Thus, in the course of one pulse-sending period, the target ranges from $\text{Д}_{\min}$ to $\text{Д}_{\max}$ are measured. Maximum counter reading N corresponds to the maximum range and is determined from the expression

$$N=\frac{\text{Д}_{\max}}{\Delta\text{Д}}, \qquad (4.2)$$

where $\Delta\text{Д}$ is the width of the range ring.

The accuracy in range coding is determined basically by the frequency stability of the clock-pulse generator and by the coincidence accuracy of the synchronizing pulse with the first clock pulse which starts the binary counter.

With the scanning of the radar coverage area, it is possible to observe several targets at different ranges. In order that the data on the range of the different targets will enter the different storage cells of the computer for further processing, a target switch and the distribution-shifter are used (Fig. 4.10). In the initial position high voltage is fed to the first line of the distributor-shifter, and to the second and third - low voltage.

With the arrival of the first detection pulse range code is read, and then over a time equal to the transit time in the delay line, the target detection pulse enters the flip-flops of the switch, returning them to the zero state.

In this case the first flip-flop forms a pulse which, in passing through the delay element (in the diagram not shown), converts the second flip-flop into state "1." Thus, the target switch is the shift register, which ensures the displacement of state "1" in turn into cells 1, 2, and 3 by pulses of the targets, which perform the role of shift pulses. The distributor-shifter makes it possible, by using the pulses of the target switches, to direct the range codes of the targets to the different storage cells of the computer.

The units of the binary number are issued from the read selector output in the form of pulses with positive polarity. These pulses are distributed in any digit by three vertical buses, each of which is a voltage divider, consisting of resistor R with high resistance, an internal diode resistor and resistor r with low resistance. The cathodes of the diodes are connected to the horizontal buses; the anodes - to the vertical buses.

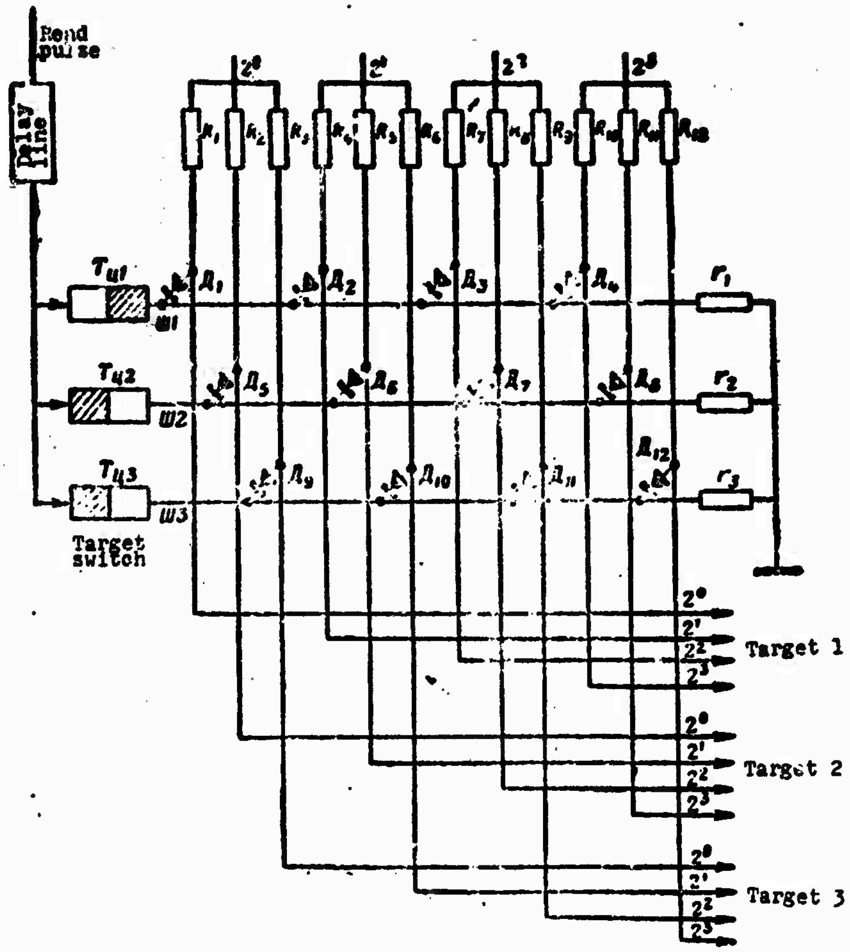


Fig. 4.10. Simplified distributor-shifter circuit.

Diodes $Д_1$-$Д_4$ of the first horizontal bus are closed by the positive voltage fed to their cathodes. Remaining diodes are open and the circuits connected to them are shunted; therefore the positive pulses of the range code can pass only through resistors R_1, R_4, R_7, and R_{10} of each digit correspondingly to the storage cell of the first target.

After the extraction of the data in the first target the positive voltage is fed to the second horizontal bus, closing the diodes connected to this bus. In this case the code of the second target is transmitted to the second storage cell, etc.

On the described principle it is possible to construct distributor-shifter for a great number of digits in the range code and for a large number of targets.

4.4. Digital Measurement of Azimuth

The transformation of angular coordinates into binary code is reduced to measuring the antenna angle of rotation (by azimuth or by the angle of elevation) and expressing it in bits. Since the reflected signal occupies a certain angle in terms of azimuth, the true direction to the target in the majority of cases is found as the arithmetic mean of the two readings: the beginning and the end of the pulse packet. Other methods of determining azimuth are also possible.

The means of determining azimuth in binary code can be realized by different methods.

If, for example, the antenna rotates in terms of azimuth with constant velocity than azimuth can be determined with the help of a circuit analogous to the digital ranging system. Actually, the angle of rotation θ at constant angular velocity can be measured indirectly - by the antenna rotation time. In this case measuring the azimuth is reduced, as in the case of ranging, to converting the time interval into a number.

Difference is that the expander (the strobe generator) is started not by the radar trigger pulse, but by the pulse corresponding to the antenna bearing at true north. The operating cycle of the circuit is determined not by the repetition period of the radar trigger pulses, but by the radar scanning time in terms of azimuth.

Such a method of measuring an azimuth can be applied only with the very rigid stabilization of the antenna scanning rate. However, under actual conditions, as a result of the inadequacy of mechanical gears, variable wind load and so forth, the rotational speed of the antenna will vary. For this reason the indirect methods of measuring azimuth practically have only limited dissemination.

Best results are obtained, if the clock pulses extracted not from the clock-pulse generator, but from the current azimuth sensor. The basic element of the current azimuth sensor is the converter of the angle of rotation into binary code. It is possible to use the most diverse converters. For example, converters with code disks, induction sensors, a magnetic drum and with other devices can be used.

One of the simplest versions of the converter is based on the use of a slotted disk. Along the periphery of the opaque disk there are slots, whose number corresponds to the discontinuity of the angle reading (Fig. 4.11), On one side of the disk there is a light source which creates a narrow beam. Each moment the beam illuminates only one slot in the disk. On the other side of the disk a photocell is installed. The photocell converts the light pulses, which penetrate through the slots in the disk while it rotates synchroneously with the radar antenna, into electric current pulses.

To fix the beginning of reading and to trip-out the counter readouts (Fig. 4.12) at the moment of passage through the conditional zero bearing ("north"), one additional slot ("reset" slot) is made in the disk and an additional photocell is installed.

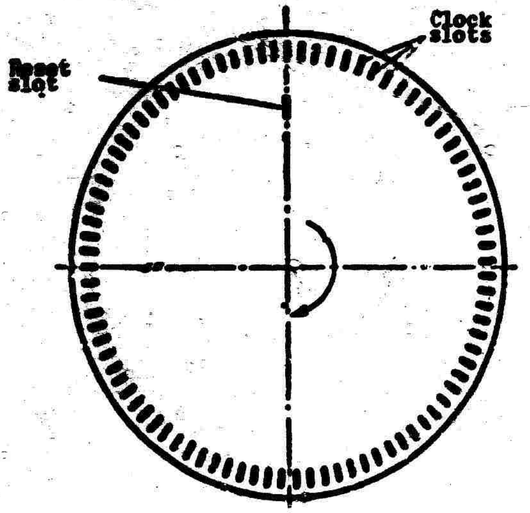


Fig. 4.11. Opaque slotted disk for the converter of the angle of rotation into a number.

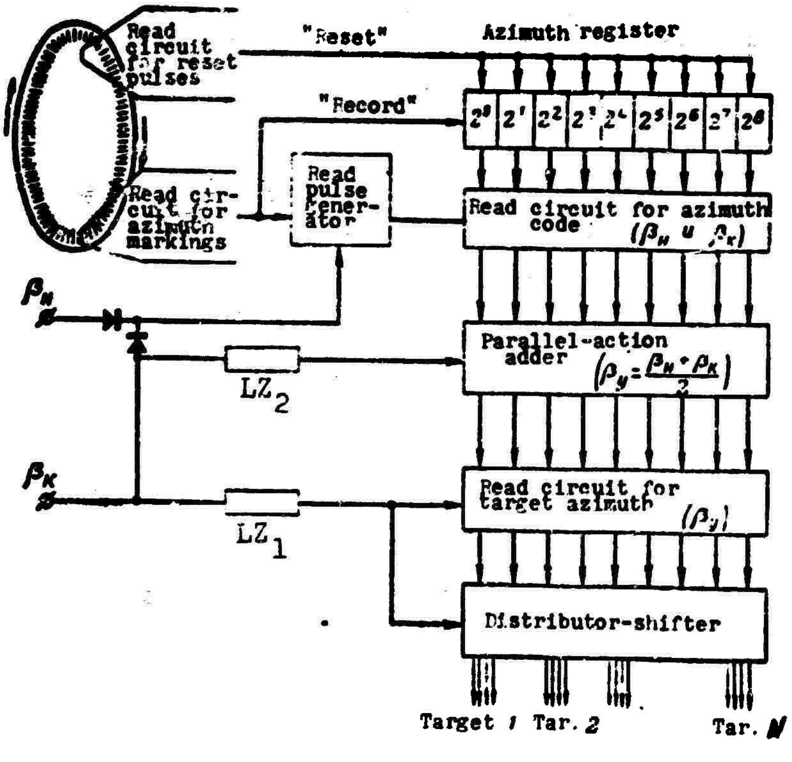

Read circuit for reset pulses
"Reset"
Azimuth register
"Record"
2^0 2^1 2^2 2^3 2^4 2^5 2^6 2^7 2^8
Read pulse generator
Read circuit for azimuth markings
Read circuit for azimuth code (β_H и β_K)
β_H
Parallel-action adder ($\beta_y = \frac{\beta_H + \beta_K}{2}$)
LZ$_2$
β_K
Read circuit for target azimuth (β_y)
LZ$_1$
Distributor-shifter
Target 1
Tar. 2
Tar. N

Fig. 4.12. Functional diagram of equipment for the automatic measuring and coding of the azimuth.

The operation of the main units of the system is analogous to the operation the corresponding units of the digital ranging system. The principal distinction is that the operating period of the azimuth measuring equipment is by approximately three orders more than the operating period of the ranging equipment.

During antenna rotation the photocell converts light pulses into electric current pulses, which are then amplified and enter the azimuth counter as azimuth marks. As a result a binary number proportional to the azimuth of the antenna will be formed in the counter.

With the pulse arrival of the beginning of packet, the read pulse generator produces a read pulse which opens the code read selectors of each digit of the counter. In this case the binary azimuthal number fixed by the counter is copied into the appropriate cells of the parallel-action adder. There the new readout of the azimuth counter, which is fixed by it at the moment the end-of-packet pulse arrives, is also copied.

The parallel-action adder sums up the binary numbers corresponding to azimuths $\beta_н$ and $\beta_к$ and divides this sum by halves.

Over time t_1, which is equal to the adder execution time in the operation of computing of the azimuth of the target β, the pulse of the end of the packet is fed to the selectors for reading the target azimuth code, for which it is delayed in delay line LZ_1 for time t_1. As a result the selectors for reading the azimuthal code are opened and the binary number of the azimuth of the target enters through the distributor-shifter to the memory unit of the computer for further processing.

The distributor-shifter and target switch circuit, which ensures the admission of the data on target azimuths to different storage cells, performs the same as in the ranging system.

Clearing (zero adjustment) the adder and preparing it for the reception of new binary numbers which characterize the azimuths of other targets is accomplished by the pulse of the end of the packet. For this the pulse is delayed in delay line LZ_2 for time $t_2 > t_1$.

Examined equipment makes it possible to measure the azimuth of targets located in the radar visual range, with change in the antenna scanning rate within sufficiently wide limits. However, with a drop in the scanning rate to zero or with a reversal, it ceases to operate normally which is its basic deficiency. The second deficiency is that the diameter of the disk should be large for the arrangement of a sufficient number of slots.

Free of the indicated deficiencies is the method for the determination of target azimuth by reading the azimuth from the code disk. The binary code of the angle of rotation is put by the photographic method on the transparent disk, which is fastened directly on the axis of the antenna (Fig. 4.13). In this case the disk is divided into a number of sectors and rings. The width of the sectors is determined by the accuracy in azimuth determination, and the number of rings - by the needed number of code digits. The outer ring corresponds to the first digit of the code, the second ring - to the second digit, etc. To each value of the rotational angle of the antenna corresponds the fully determined, intermittent combination of light and dark sections. On one side of the disk there are point light sources, and on the other - a screen with a narrow slit. The screen transmits light to miniature photocells, whose number equals the number of digits (Fig. 4.14). The light beam passing through the disk causes the appearance of current pulses in the circuit of the photocell. After amplification the pulses enter the read selectors, fixing the code units in the appropriate digits. If in the other digits the luminous flux is covered by the shading in the disk, then there are no pulses at the output of the amplifiers for these digits. Zeros are fixed in the digits.

Such a system can measure the angle of rotation of the disk regardless of the fact that it rotates, is stable or oscillates. Target azimuth in this case is determined by the following manner.

The beginning-of-packet pulse which enters from the pre-selector opens the read selectors (I-0-I-4), at whose outputs a code combination of zeros and units is created. This combination corresponds to the antenna position at a given instant. The fixed binary number corresponding to azimuth $\beta_{\text{н}}$ of the target reenters the parallel-action adder.

Upon the termination of the pulse packet the new combination of zeros and units, corresponding to the new position of the disk enters the adder. The adder calculates azimuth $\beta_{\text{ц}}$ of the target. In other respects the circuit in question operates just as the circuit shown in Fig. 4.12.

However, a substantial deficiency is inherent in the circuit in Fig. 4.13. It is caused by the appearance of large errors during the reading of the code, when the reading occurs at the boundary of the code combinations of two adjacent angles. For example at the boundaries of numbers 7 and 8, 23 and 24 there is an exchange of units and zeros immediately in four digits, but at the boundaries of numbers 15 and 16, 31 and 0 - even in five digits. If luminous fluxes are set-up with insufficient accuracy then different number combinations which do not reflect the real azimuth code are possible at such borders.

The method of coordinating the moment of the reading with the transit process time, is unsuitable here, since there are no clock pulses used in the previous circuit. To eliminate the errors in the readings not an ordinary, but cyclic binary code (the Gray code, Fig. 4.13b) is placed on the disk. In all cases this code gives an error which does not exceed the unit of the low-order digit.

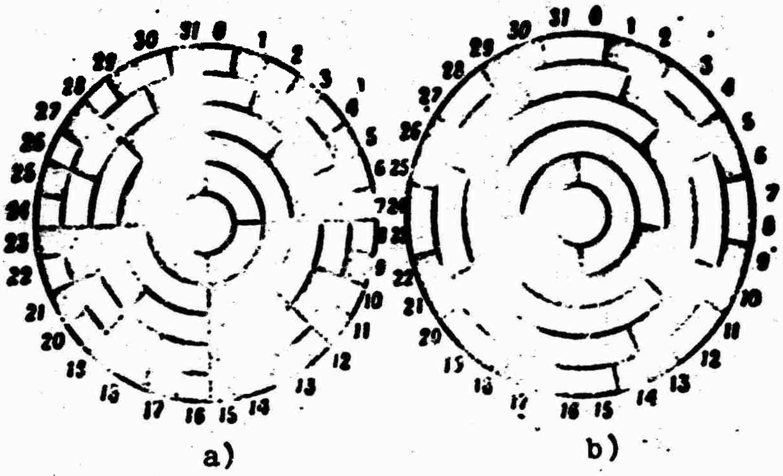


Fig. 4.13. Code disks with the applied binary a) and cyclic b) codes.

For comparison of the different codes Table 4.3 gives decimal, binary and cyclic codes. The section of cyclic code combination is shown here for the sixteen positions of the disk. Light places correspond to the transparent disk sections which give the unit during the reading. Places with xxx correspond to zero. For convenience the digits are arranged not in the form of a ring, but in the form of direct bands.

It is apparent from the cyclic code examination that all the digits are equivalent, and during the transition from one binary number to another there is a change only in one digit. Consequently, it is feasible to read either one or the other nearest number on the boundary of two digits. In this case an error in the reading will comprise only a small fraction of one division, i.e., part of the unit of the low-order digit.

Table 4.3.

Decimal code	Binary code	Cyclic code	Digits			
			4	3	2	1
0	0000	0000				
1	0001	0001				XXX
2	0010	0011			XXX	XXX
3	0011	0010			XXX	
4	0100	0110		XXX	XXX	
5	0101	0111		XXX	XXX	XXX
6	0110	0101		XXX		XXX
7	0111	0100		XXX		
8	1000	1100	XXX	XXX		
9	1001	1101	XXX	XXX		XXX
10	1010	1111	XXX	XXX	XXX	XXX
11	1011	1110	XXX	XXX	XXX	
12	1100	1010	XXX		XXX	
13	1101	1011	XXX		XXX	XXX
14	1110	1001	XXX			XXX
15	1111	1000	XXX			

Cyclic code reads itself by the circuit in Fig. 4.14. In this case the Gray parallel code is obtained. The number obtained in cyclic code must be fed to the adder to compute the target azimuth in the form of a binary code. For this purpose translator of the code from a cyclic to a binary code is included in the circuit. For simplicity in the code translator circuit, the parallel cyclic code is converted into a parallel binary code in two stages. First the parallel cyclic code is converted into a sequential cyclic code, then - into a binary code. To convert the parallel cyclic code into a sequential code shift registers are used. One of the versions of these registers was examined in the digital computer for radar signal detection. The parallel code is input simultaneously into all the cells of register, and then the shift pulses alternately move the recorded code forward into the last cell of the register, whence the sequential cyclic code is also removed.

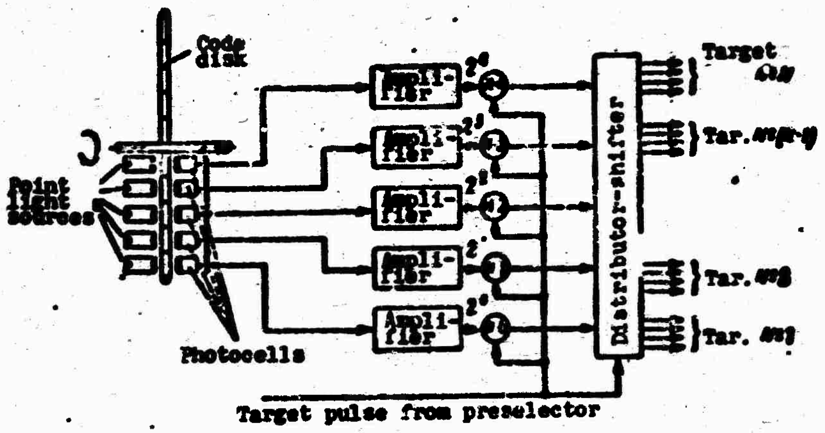


Fig. 4.14. Functional circuit of the equipment for the extraction of the azimuth code using the code disk. [И = I].

4.5. Algorithm for the Primary Processing of Radar Information

Usually the device used for the primary processing of radar information is a specialized electronic digital computer, whose discharged functions can be divided into two groups: the determination of the plane coordinates of the targets and their identification; and the determination of the altitude and characteristics of the targets.

The first group includes: obtaining information from circular scan radar and identification equipment; the extraction of target signal from the received information; determination of the range and azimuth of the target; typing the identification signals obtained from the identification equipment to the target coordinates; the temporary storage of the radar information; and the conversion of the information into a form convenient for transmission of the data to higher headquarters over the lines of communication.

The second group includes: the reception from the control post of request instructions (target designations) concerning the need for the altitude measurement or for the determination of additional target characteristics; the generation of control signals which ensure the turning of the altimeter antenna to the assigned azimuth; providing the operators with visual data on the characteristics of specific targets; and the preparation of data for the operator for decision-making on the transmission of the information to the command post.

Radar information from the output of the equipment which issues data in binary code is usually transmitted for recording onto a magnetic drum (MD) which rotates at high speed. A magnetic drum is like the delay lines, in which the information (obtained at different ranges) is stored. Each track of the drum has a recording head and a reproducing head. The rotational speed of the magnetic drum and the transmitting frequency of the radar transmitter should be precisely synchronized in this case.

At a specific moment in time the reproducing heads begin to read the recorded target signals for a given range. The number of recorded pulses, with the help of storage circuits and comparison circuits, is compared with the preset number, which is useful information. The useful information from the signals received is separated by the above-described methods, and, as soon as the recorded pulse packets are identified with the targets their characteristics (range, azimuth, approximate group composition, etc.) are automatically determined and presented at the output of the machine.

Identification information, as a rule, is processed in a separate channel of the computer and is accompanied by special tag in the issued report.

With a complex air situation the information on some targets can be stored in the computer for several seconds before it enters the line of communication. Thus, the measurement of the time interval between the moment the last pulse of the packet is received and the moment the information is transmitted to the output device is provided for in the computer. If the lag of the information exceeds a set magnitude, the information is considered obsolete and is erased.

All the information about the target - target number, azimuth, range, pulse width of the packet, storage time, identification tag and synchronizing pulses - is concentrated at the output device for the computer information. Other target characteristics, as a rule, are determined and transmitted by special request from the control post.

Let us examine the algorithm for gathering and processing the information. This algorithm is realizable in the computer in the primary processing stage.

As a rule, the algorithms for a different kind of problems are presented in the form of a structural (block-diagram) or a operator recording. In this case a principle is used, on whose basis the process of solving any problem can be divided in a series of independent arithmetic and logical stages called operators. Regardless of the conditions arithmetic operators fulfill the specified assigned commands always equally. The logical operators fulfill the next instruction depending upon some condition.

Thus, for instance, the logical operator which checks the targets for the "ours-foreign" tag, leads to the execution of different commands depending on the presence of one or the other tag. If the target has an "ours" tag, it is eliminated from distribution; if a "foreign" tag - it is left in the distribution. Before constructing the algorithm, analysis is made of the problem, and in accordance with its outcome a resolving method is selected.

Every operator can be a certain group of commands for a given stage of solving the problem, and it can also consist of a single command. The automatic problem solving process on the computer, besides the arithmetic and logical stages, also contains some others, for example: the transfer of numerical material from punchcards, punched tapes and magnetic tapes to the computer storage and vice versa (the transfer operators correspond to these stages); and machine halting to which the halt operator corresponds.

Each operator is designated by its symbol (letter): the arithmetic - by the letter A; the logical - by the letter R; the transfer operators - by the P; and the machine halt operator - by the letter Ya.

Figure 4.15 gives the block-diagram of the algorithm for the primary information processing by the computer.

In the structural record of the algorithm the problem resolving process is separated into stages, each of which is illustrated by a rectangle (block) with a record of the stage content. Blocks are connected by arrows, indicating from which block and to where the control is passed on in the course of solving the problem.

The first stage - the conversion of the radar pulses from radar code into the binary digital code of the computer - is expressed by transfer operator P_1.

The second stage is the determination of presence in this signal of a sufficient minimum of useful information by means of comparison with the established criterion. If the signal contains the necessary minimum of information, permission for transition is given to the following operation. If there is

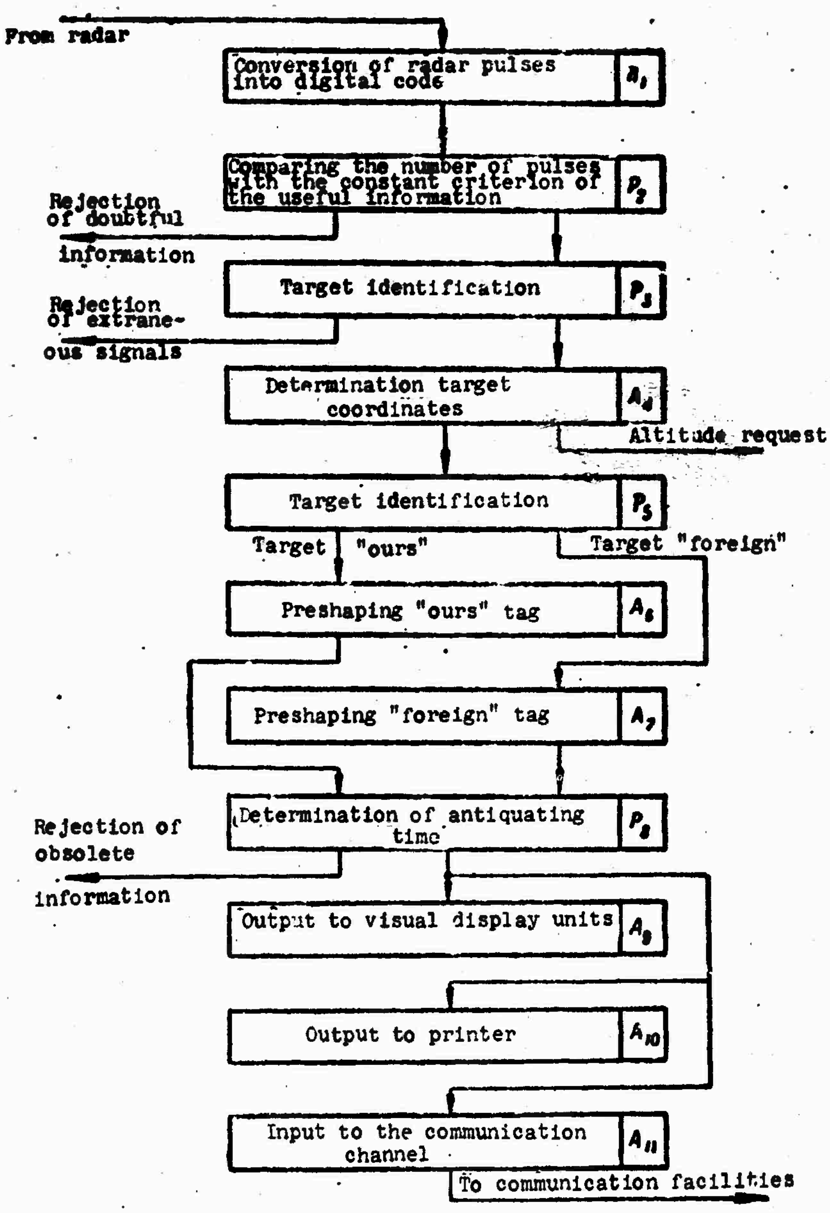

From radar
Conversion of radar pulses into digital code
A_1
Comparing the number of pulses with the constant criterion of the useful information
P_2
Rejection of doubtful information
Target identification
P_3
Rejection of extraneous signals
Determination target coordinates
A_4
Altitude request
Target identification
P_5
Target "ours"
Target "foreign"
Preshaping "ours" tag
A_6
Preshaping "foreign" tag
A_7
Determination of antiquating time
P_8
Rejection of obsolete information
Output to visual display units
A_9
Output to printer
A_{10}
Input to the communication channel
A_{11}
To communication facilities

Fig. 4.15 Block-diagram of the algorithm of primary information processing. Designations: П = P, P = R.

no necessary minimum, the signal is rejected. Logical operator R_2 corresponds to the second stage.

In the third stage the target is identified by comparing the signal with the different type characteristics (speed, magnitude, density of the signal, etc.) of the normalized (averaged) target. If it is established that the signal corresponds to the characteristics of aircraft or missiles, proceed to the following stage; if conformity is not established, then signal is rejected. Logical operator R_3 corresponds to the third stage.

In the fourth stage, which is expressed by arithmetic operator A_1, target coordinates Д, β are computed and a request for determining the altitude of the target is transmitted. After this they proceed to the next stage.

The fifth stage, which is expressed by logical operator R_5, identifies the target. If the identification equipment takes the "foreign" tag together with the signal from the target, control is passed on to the seventh stage, if the "ours" tag is taken - to the sixth stage.

The sixth stage (arithmetic operator A_6) adds the "ours" tag to the signal from the target, after which control is passed on to the eighth operator. The seventh arithmetic operator A_7 completes similar action, but with the "foreign" tag.

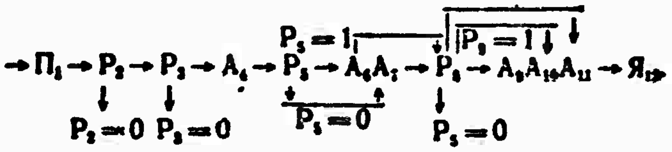
The eighth stage (logical operator R_8), by comparison with the criterion of antiquating the information with time, checks whether this information has become obsolete or not. If the calculated time of antiquating exceeds the permissible limit, the information is rejected. If the time of antiquating is less than the permissible limit, control is passed on in parallel to the three arithmetic operators A_9, A_{10}, and A_{11}.

A_9 - the arithmetic operator of information output to the visual display equipment.

A_{10} - the arithmetic operator of information output to the printer.

A_{11} - the arithmetic operator of information input to the communication channel for transmission to the headquarters of higher authority.

At this point primary information processing ends. The logic circuit of the examined algorithm for primary information processing can also be written in operator form. In this case the logic circuit of the algorithm is illustrated as the alternation of operators, each of which is furnished with a digit, which indicates the ordinal number of the operator:



Designations: П = P, P = R, and Я = Ya.

The arrows show the direction in which the control proceeds from one operator to another.

The logical operators, unlike the others, can (as already mentioned) fulfill one of two commands. For example, logical operator R_5 for target identification is expressed this way:

$$R_5 = \begin{cases} 1, \text{ if the target is "ours";} \\ 0, \text{ if the target is "foreign."} \end{cases}$$

For convenience the logic circuit of the algorithm is sometimes written in one string; in this case the arrows between

the neighboring operators are omitted. If during such a record the operator symbols stand in queue, then this means that the operator to the right gets control from its neighboring operator to the left. If the operator to the right does not gain control from the operator to the left, then a semicolon is put between them.

In a case where control is passed on not to the operator to the right, but to any other, the transfer of control is designated by an arrow. The arrow which corresponds to condition "1" of the logical operator is drawn above the operator string and the arrow corresponding to condition "0" - below the string. If the transfer of control is not bound by any conditions; the arrow is drawn abitrarily.

The above-given logic circuit of the algorithm for primary processing in accordance with the indicated rules will be written in the following manner:

$$\Pi_1 P_2 P_3 A_4 \overset{\overset{7}{\uparrow}}{P_5} \overset{\overset{8}{\uparrow}}{A_6} ; \overset{\overset{5}{\downarrow}}{A_7} \underset{\underset{10,\,11}{\downarrow}}{\overset{\overset{6}{\downarrow}}{P_8}} A_9 ; \overset{\overset{8}{\downarrow}}{A_{10}} ; \overset{\overset{8}{\downarrow}}{A_{11}} \text{Я}_{12}.$$

Designations: П = P, Р = R, and Я = Ya.

The solution to the problem on this algorithm begins from the operation of the operator to the extreme left, after whose execution of the actions follows the execution of the actions of the operators in order. When the queue reaches the logical operator, then upon the execution of logical condition the actions prescribed by the next operator to the right are performed, and with failure - the actions are performed by the operator, to whom the arrow from the logical operator points. The algorithm action is finished by the halt operator.

4.6. Equipment for the Semi-Automatic Extraction of Information

Along with equipment for the automatic extraction of radar information at the present time being widely used is equipment for semi-automatic extraction from radar linked with the ACS.

The essence of the semi-automatic extraction of radar information consists of the following. Blips from targets are reflected on screens of the scopes are observed visually by the operator. The operator selects from all the blips those which must be taken for tracking, and with the help of special equipment the coordinates of these blips are extracted.

The semi-automatic extraction equipment can be coupled both with the conventional different types of scopes and with special indicators. With the use of conventional scopes the latter are equipped with opto-mechanical extractors, and with the special scopes of the ACS they are linked by electron-optic extractors.

The opto-mechanical method of extracting information consists of the fact that the light spot (light marker) whose position varies on the screen scintillates with the help of the extractor on the screen of the cathode-ray tube. Operator, by turning the lever, mechanically combines the light marker with the target blip. Since the extractor lever is connected to the transmitters of coordinates on the X and Y axes; with the coincidence of light marker with the target blip on these transmitters, mastered will be the values of the target coordinates ($X_{\text{ц}} - X_{\text{марк}}$ and $Y_{\text{ц}} - Y_{\text{марк}}$).

In practice the electron-optic method, with which an electronic marker whose position is determined by the values of the voltages extracted from the transmitters of the extractor

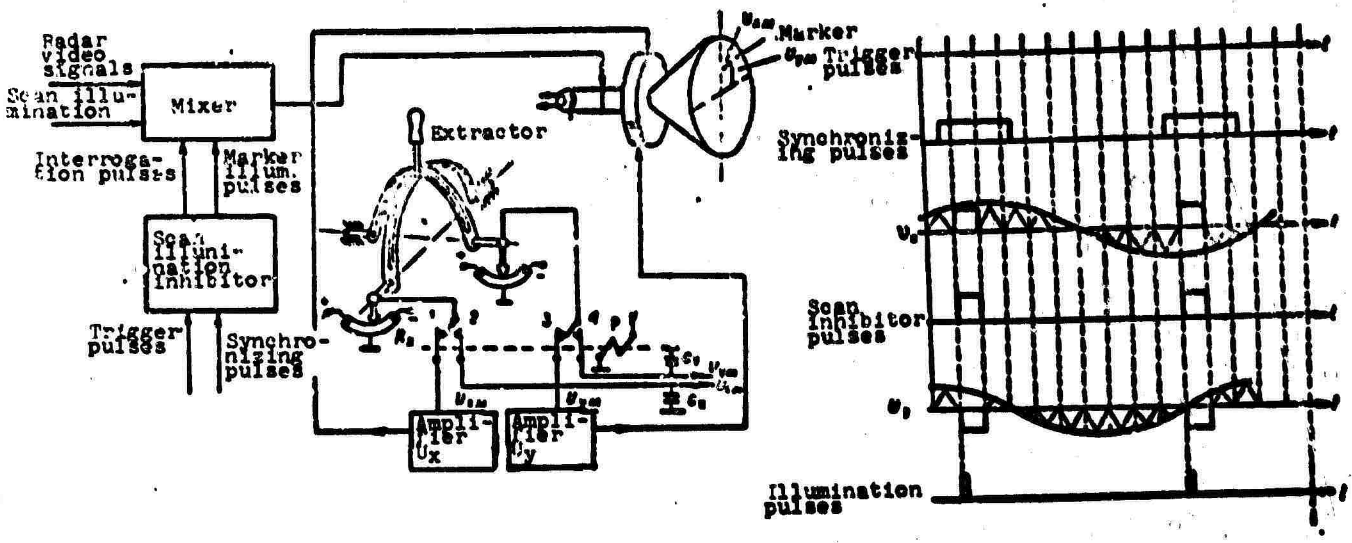


Fig. 4.16. Functional circuit of electron-optic extraction and curves of the voltages which are at work in the circuit.

equipment is reproduced on the cathode-ray tube screen, has been largely disseminated. The scopes simultaneously scintillate the primary radar situation being supplied from the output of the radar receiver R.S. and the so-called secondary situation which is the blips of the electronic markers, various map symbols and digits.

The electronic marker is fed to the scope screen (Fig. 4.16). The position of the marker is determined by voltages U_{xm} and U_{ym}, which are extracted from potentiometers R_X and R_Y of the extractor. The brushes of the potentiometers are moved by the operator with the help of mechanical linkage. As a result of this the marker moves over the screen. At the moment the marker coincides with the target blip, the extraction button is pressed and relay P triggers. The contacts of the relay switch on the coordinate potentiometers to accumulators C_X and C_Y, on which voltages U_{xm} and U_{ym}. Below, these voltages can either be transmitted directly to the display units or be precoded, and then transmitted over communication channels.

To reproduce the blip of the marker it is necessary for a certain time to discontinue admitting to cathode-ray tube deflection coils the time-base sawtooth voltages and at this moment to supply the constant voltages of the marker. The frequency at which it is necessary to scintillate marker should be such that on one hand the marker would be observed as nonscintillating point, and, on the other, that the loss of information because of scan discontinuity would be minimal. It has been established that these requirements are satisfied by a marker scintillation frequency of 15-16 Hz. Scan is interrupted by a special inhibitor, which is synchronized from the general timing mechanism. Besides scan discontinuity, it is necessary to feed the marker illumination to the control electrode of the cathode-ray tube.

Besides the described opto-mechanical and electron-optic methods of the extracting radar information, there is also the electronic method, which has many diversities. The electronic semi-automatic extractors are referred to as a different kind of "pen styluses" (light pen stylus, beam pen stylus, electropen stylus, piezopen stylus, magnetic pen stylus, etc.).

CHAPTER 5

SECONDARY PROCESSING OF RADAR INFORMATION

5.1. Definition and Makeup of Secondary Processing

Aerial target information obtained as a result of the primary processing of radar information can contain those or other errors caused by the random nature of the signal reflected from the target.

Because of the incomplete breakdown of false information in primary processing equipment various kinds of interference signals enter together with the target signals the input of the equipment for further data processing. The totality of the information on the "true target" coordinates at the output of the primary processing equipment is acceptably called the target blip. Correspondingly, the total set of "false target" coordinates is called a false blip.

A single target blip only approximately reflects the true position of the target in the visual range of the radar at the moment of location. By it alone it is still not possible to make a reliable decision about target detection, much less to judge such target movement parameters as speed, acceleration, course, etc.

In connection with this, it becomes obvious that primary data processing based on analysis of the reflected signals of a target within the limits of one radar scanning period does not give exhaustive information on targets within visual range. It is only the initial stage of data extraction. In order to make the correct decision on the presence of a target and to determine its movement parameters, it is necessary to analyze the information obtained per several scanning periods. The operator, who observes the screen of the circular-scan scope uses the same approach.

If at any point on the screen a single blip appeared, the operator records it as a possible target. If in the next scan the blip appeared again and furthermore moved a certain distance, then there is already a basis for making a decision on target detection. Naturally, by using the blips obtained after three sequential scans, the probability of correct target detection will be even greater. At the same time it is possible to determine the heading and speed of target movement.

Thus, during the observation of signals on the radar scope, it is possible to note that the target blips are moving and dr [illegible] flight trajectory, repeating with certain accuracy the movement of the target in the area. Due to this the target blips mainly retain those regularities which are characteristic of the target itself.

For real moving objects (targets), characteristic is the fact that between their previous and subsequent position there is a connection caused by the inertial properties of real targets and by their ultimate maneuverability capabilities. Therefore, it is possible to assertain that over a certain, rather short time the subsequent blip should be located in vicinity of the previous one. Moreover, it is possible with definite accuracy to guess the coordinates of the next blip, if we first compute the value and heading of the target's flight speed. Similar regularities develop more strongly the smaller the interval between the previous and subsequent blips. With an increased interval such a connection, as a rule, weakens and can disappear entirely.

The most probable reasons for the appearance of false blips are caused by purely random factors. In view of this the laws governing the appearance and disposition of blips on the radar scope prove to be different. They are developed first of all when there is no connection between blips from scan to scan. False blips appear chaotically, in different places on the scope screen, while the target blips are situated (with a certain spread due to errors in measuring) along the trajectory of target movement.

By observing position of the blips on the scope screen from one scan to another, the radar operator distinguishes the false target blips from real ones.

The operations performed by the operator can principally be formalized, and their execution can be laid upon a specialized electronic digital (or analog) computer. In this case in the semiautomatic processing system, the automations will be subject only to the operation of determining the parameters of target movement and producing of predicted coordinates. The other processing tasks are performed in the semiautomatic system by the operator visually or with the help of mechanized equipment (extractors), which make it possible to increase the rate and the accuracy of visual processing.

Unlike the semiautomatic system an automatic processing system solves the problems of processing target trajectories completely using electronic digital computers. In this case the computer, whose storage capacity and operating speed should be sufficient, analyzes the change in the coordinates of the targets and establishes the laws governing their change from scan to scan for each target individually. On the basis of the given analysis, it is possible to reject the false targets because of the absence of regularity in their change.

Automatic or semiautomatic processing of information obtained after several radar scans for the purpose of detecting and continuously tracking the trajectories of the targets is called the secondary processing of radar information.

It includes the following operations:

- the determination of the target movement parameters (course, speed, acceleration, etc.) by the data obtained after several radar scans;

- the isolation of the area of space in which with certain probability the appearance of a blip in the next sweep (blip extrapolation) is expected;

- the comparison of the extrapolated coordinates with the new ones and tying the new blip to the target trajectory (continuation of the trajectory).

Secondary processing equipment separates the trajectories of the moving targets from the individual blips given out by the primary processing equipment.

Thus, if during primary processing the useful information is separated from the signal and noise mixture on the basis of the statistical distinction in the structure of the signal and the noise, then the secondary processing, using the distinctions in the laws governing the appearance of false blips and target blips, should ensure the separation of the trajectories of the moving targets.

5.2. Construction of the Trajectory of Target Movement

The basic data which determine the trajectory of target movement are the space coordinates of the target blips whose change corresponds to the law of target movement in space.

In general. target movement can be described by the coordinates of its center of mass, which are random functions of time. The precise determination of the pattern of these functions requires analysis of the utilization tactics of the appropriate facilities and their maneuverability capabilities.

The trajectory of target movement depends upon many factors and conditions, such as the type of target, flight altitude, speed, maneuverability capabilities, etc. Furthermore, the target trajectory is influenced by a whole series of random factors (disturbances), under which are implied all the reasons which misrepresent trajectory or which hinder its detection and reproduction. Those pertaining here, for example, are the random oscillations of the target around a kinematic trajectory, which are caused by the effect of the random perturbations of the medium, by errors in the target control system, by instrument errors in measuring the coordinates by radar, by primary processing errors, by false blips, by blips from ground features, as well as by man-made disturbances. Let us briefly examine each of the component disturbances.

The random purturbations of the target around an assigned trajectory are a process with the normal probability distribution law. The mathematical expectation of such a process equals zero, since on the average the target is held to the assigned trajectory. The random deviations are characterized by dispersion and the correlation function which determines the connection of these divergences in adjacent radar scan. Correlation function is found experimentally for each type of target and for each concrete control system.

Only the random component is considered in the errors of measuring the coordinates; systematic component is not considered, since it can be compensated for. Random errors are subjected to normal distribution law of distribution. For circular-scan radar having a scanning period of several seconds, random errors in coordinate measurement can be considered uncorrelated in the neighboring scans.

False blips appear randomly and independently within the limits of the entire radar visual range. It assumes that the false blips are distributed evenly in terms of time; with average density ρ (blips/s) it is therefore possible to compute the density of the false blips per unit of area. By knowing the total number of false blips which appear in the visual range during period T_0 which equals ρT_0, we will get the density of the false blips

$$\nu = \frac{\rho T_0}{n}, \qquad (5.1)$$

where n - the number of assigned units of area.

Thus, the statistically false blips can be characterized either by average density with time ρ (blips/s), or by density ν per the unit of visual range area.

Target-reflected pulses can fade in certain cases due to fluctuations. This fading can be described by the Poisson law.

The enumerated and certain other factors compel one to attribute the movement of the targets to the category of processes with parameters which randomly change with time. Evidently, for a statistical description of such processes it is necessary to know the probability distribution laws of the parameters which determine these processes. However, it is possible to obtain such laws practically. Therefore, it is necessary to set some hypotheses concerning the statistical characteristics of the processed signals, i.e., to proceed from a more or less likely statistical model of the target's movement.

The selection of a specific model of target movement depends on what kind of targets the secondary processing equipment will specifically have to deal with. Thus, for example, if system is intended for processing the trajectories of ballistic missiles and earth satellites, then the models of their movement can be represented by the equations of curves of the second order (ellipse, parabola, circle). If the system is intended for processing the trajectories of aerodynamic facilities, for example, such as aircraft and winged missiles, then the models of their movement are a total set of sectors with rectilinear and uniform movement and a total set of maneuvering sectors. A polynomial movement model can be taken as the basis for this type of equipment. It is based on representing the process of change in target coordinates in a limited observation zone in the form of a polynomial of degree n relative to time:

$$Y(b;\ t)=\sum_{i=0}^{n} b_i t^i = b_0 + b_1 t + b_2 t^2 + \ldots, + b_n t^n, \tag{5.2}$$

where b_i - a coefficient which determines the trajectory parameters upon which these or other limitations are imposed.

On the strength of the nature of the movement of aerodynamic targets (random alternation of the sectors of rectilinear uniform flight and maneuvering), three basic requirements for automatic tracking systems are formulated:

- in the sectors for straight flight and in maneuvering sectors, the hypotheses about the nature of coordinate change with time must be different;

- in the straight flight sector the change in coordinates with time is more simple to describe by polynomials of the first degree, i.e., to make a hypothesis about the rectilinear, uniform movement of the target:

$$\begin{aligned} x(t) &= x_0 + V_x t; \\ y(t) &= y_0 + V_y t; \\ H(t) &= H_0 + V_H t; \end{aligned} \tag{5.3}$$

- in the maneuvering sector the process of change in coordinates with time is best of all described by polynomials of the second degree:

$$\begin{aligned} x(t) &= x_0 + V_x t + \frac{a_x}{2} t^2; \\ y(t) &= y_0 + V_y t + \frac{a_y}{2} t^2; \\ H(t) &= H_0 + V_H t + \frac{a_H}{2} t^2, \end{aligned} \tag{5.4}$$

where a_x, a_y, a_H - accelerations at each coordinate.

Representing the coordinates of a maneuvering target by a polynomial of a degree higher than the second does not give considerable advantages in the accuracy of estimating the trajectory parameters.

Thus, the trajectory of target movement is represented in the form of a sequence of polynomial sectors with different coefficients and degrees of polynomials. The processing system should therefore be reconstructed in accordance with the nature of the movement of every target.

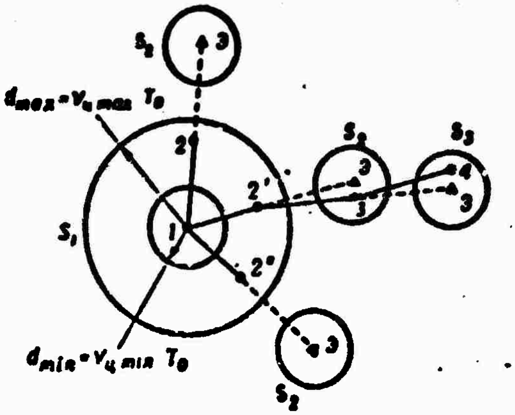
The polynomial model of target movement has a substantial disadvantage, which consists of the impossibility of accounting for the unforseen (unexpected) maneuvering of the target.

5.3. Automatic Detection (Automatic Lock-On) of Trajectories

The process of secondary processing is broken down into two separate stages: trajectory detection and trajectory tracking.

The automatic detection (automatic lock-on) of trajectories is the beginning moment of secondary processing. One of the feasible methods of automatically locking-on to the trajectory of a target is presented in Fig. 5.1.

Fig. 5.1. For the explanation of the automatic lock-on to the trajectory of a target.



Suppose single blip 1 from the target appeared. It is taken as the initial blip of the trajectory. In the following scan the second blip belonging to that trajectory should be searched for in a certain area included within the ring with area

$$S_1=\pi T_0^2(v_{u\,max}^2-v_{u\,min}^2), \tag{5.5}$$

where T_0 - radar scanning period;

$V_{u\,min}$ and $V_{u\,max}$ - possible minimum and maximum target speeds.

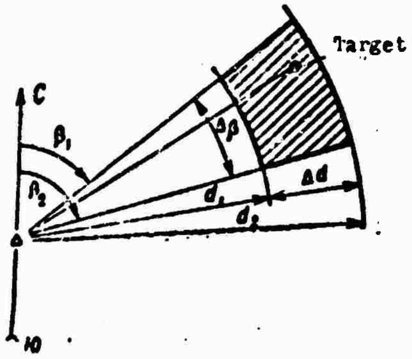
Not one, but several blips can fall into area S_1, and each of them should be counted as a possible continuation of the supposed trajectory. The speed and the heading of the movement of each of the supposed targets is calculated using the two blips, and then the position of the blip for the following (the third) scan is extrapolated. Circular areas S_2 are formed around extrapolated blips 3.

If a blip fell into any area S_2 in the third scan, it is considered as belonging to the detected trajectory, the trajectory is extended and the blip is passed on for tracking.

The examined method for automatic lock-on results from analysis of radar operator's actions during visual trajectory detection. All the automatic lock-on operations can be formalized and their execution be laid on computers. The operations which are performed in the automatic lock-on process are reduced to the extrapolation of the coordinates, to smoothing them and to strobing the blips. In this case the coordinates are usually extrapolated on the strength the hypothesis about the uniform and rectilinear movement of the targets, which is caused by the little observation time. The simplest shape in the form of areas limited by azimuth and range is given to the automatic lock-on zones (Fig. 5.2). The accepted unit for measuring the area of the zone is the area of the elementary sector:

$$S_0 = \Delta d \Delta \beta. \tag{5.6}$$

Fig. 5.2. Trajectory automatic lock-on zone.



If we take the distribution of false blips in the visual range as being uniform, then average number of blips within the limits of the elementary area is

$$n_0 = \frac{N_0}{N_3}, \tag{5.7}$$

where $N_3 = \frac{2\pi D_{max}}{\Delta d \Delta \beta}$ - the number of elementary areas in the circular scan zone of the radar.

It is acceptable to evaluate the quality of automatic lock-on:

- first, by the average number of false trajectories being transmitted for tracking per unit of time;

- second, by the reliability of the automatic lock-on (by the probability of detecting the true trajectory);

- third, by the required operating speed and by the storage capacity of the computer, which realizes the automatic lock-on algorithm.

The problem of automatic lock-on in essence is the problem of checking the statistical hypotheses. By analogy with the detection of a target against the background of interference, during primary data processing let us take hypothesis H_0 concerning the fact that the obtained sample $x_1, x_2, \ldots, x_n$ of connected blips belongs to a false trajectory, as well as hypothesis H_1 that this sample belongs to the true trajectory.

With the automatic lock-on of the trajectories it is advantageous to reduce the number of observations in order that lock-on is made as quickly as possible. From this viewpoint, during the composition of the automatic lock-on algorithm it is convenient to use the sequential analysis criterion, whose peculiarity is that there are many stages in the statistical experiment for the selection between hypotheses H_1 and H_0. The decision on continuation or curtailment of the experiment in this stage depends on the outcome of the previous stage.

Using the sequential analysis criterion for the likelihood function of the hypotheses,

$$\begin{aligned} W(x_1, x_2, \ldots, x_n/H_0) &= W(x_1/H_0)\,W(X_2/H_0)\ldots W(X_n/H_0); \\ W(x_1, x_2, \ldots, x_n/H_1) &= W(x_1/H_1)\,W(X_2/H_1)\ldots W(X_n/H_1), \end{aligned} \tag{5.8}$$

and the likelihood ratio is

$$\lambda(x_1, x_2, \ldots, x_n) = \prod_{i=1}^{n} \mu_i, \tag{5.9}$$

where $\mu_i = \dfrac{W(x_i/H_1)}{W(x_i/H_0)}$ - the quotient of the likelihood ratio, calculated using the results of the i-th test.

If after the next test likelihood ratio (5.9) exceeded upper threshold B, the decision concerning the fact that hypothesis H_1 is corrected, is accepted, but if this ratio is less than lower threshold A, then the decision on the validity of hypothesis H_0 is accepted. If $A < \Lambda(x_1, x_2, 111, x_n) < B$, then tests continue.

The values of the upper and lower thresholds can be determined by the formulas:

$$A = \frac{P_{\text{пц}}}{1 - P_{\text{лт}}}; \quad B = \frac{1 - P_{\text{пц}}}{P_{\text{лт}}}. \tag{5.10}$$

The application of the sequential analysis criterion makes it possible to reduce the number of observations in making a decision about trajectory detection, and this increases the automatic lock-on range and reduces the volume of required storage in secondary processing computers. However, the optimal sequential algorithms of automatic lock-on, which are based on the creation of the likelihood ratio, prove to be too cumbersome, therefore it is more expedient to construct algorithms on the basis of detection criteria "l/m" and "n/m" examined in Chapter 3. In this case values l and m (for criteria "l/m" and "m/m" respectively) are taken as being equal to 2 or 3, which it is completely sufficient for solving the practical problems of trajectory detection.

5.4. Extrapolation of a Target Trajectory and its Smoothing

As has already been shown, the automatic lock-on to trajectories is composed of the extrapolation of blips, smoothing them and strobing them. Let us examine these operations in more detail.

The extrapolation process consists of the fact that the coordinates of the next blip are calculated by using the coordinates of the previously obtained blips. This problem is complicated because of the fact that the blip coordinates being utilized for prediction contain errors. Because of this the results of the extrapolation also have errors, in certain cases exceeding the errors of the initial data.

Extrapolation requires a knowledge of the laws governing the movement of targets, on the basis of which the trajectory runs.

Many methods of extrapolating target coordinates exist.

In the simplest case for computing the coordinates of an extrapolated blip, it is necessary to find speed V_n in the i-th scan.

The component for velocity V_n for coordinate x is

$$V_{nx}=\frac{x_n-x_{n-1}}{T_0}, \qquad (5.11)$$

where T_0 - the scanning period.

The coordinate of the extrapolated blip is

$$x_{\text{э}}=x_n+V_{nx}T_0 \qquad (5.12)$$

Similar calculation is made for coordinate y.

However, such a prediction of coordinates on two blips in the n-th and (n-1)-th scans gives considerable errors due to errors in measuring the speed, therefore the predicted coordinates are made more precise by averaging the speed in each subsequent scan. Averaged speed $\bar{V}_n$ with consideration for measured value V_n in the n-th scan is usually determined by the formula

$$\bar{V}_n=\bar{V}_{n-1}+\frac{V_n-\bar{V}_{n-1}}{n-1}, \qquad (5.13)$$

where $\bar{V}_{n-1}$ - the averaged speed for the previous scans without the n-th blip.

Such a prediction method is good for the case of the uniform and rectilinear movement of the target (i.e., when the coordinates of target movement are described by a polynomial of the first degree).

If the target maneuvers (i.e., its coordinates change in accordance with a polynomial of the second degree), then coordinate

$$x_{\text{э}}=x_n+V_{nx}T_0+\frac{\bar{a}_{nx}}{2}T_0 \qquad (5.14)$$

where $\bar{a}_{nx}$ - the averaged acceleration of coordinate x with consideration of the n-th blip.

In general the coordinate of the extrapolated blip is described by the algorithm

$$x_n = \sum_{i=1}^{n} \eta_{эi} x_i \qquad (5.15)$$

where n - the number of blips utilized for the prediction (the blip of the last scan is taken as a blip with the subscript n = 1);

$\eta_{э\,i}$ - the weight coefficient of the i-th blip.

Weight coefficients η can be determined by the formulas:

- for a nonmaneuvering target

$$\eta_{i(нм)} = \frac{6i-2n-4}{n(n-1)}; \qquad (5.15)$$

- for a maneuvering target

$$\eta_{i(м)} = \frac{3}{n(n-1)(n-2)}[(n+2)(n+3)-2i(4n+7)+10i^2]. \qquad (5.17)$$

Let us determine these coefficients for the nonmaneuvering target during extrapolation in three sequential scans (n = 3):

$$\eta_{э3(нм)} = \frac{6\cdot3-2\cdot3-4}{3\cdot2} = \frac{4}{3};$$

$$\eta_{э2(нм)} = \frac{6\cdot2-2\cdot3-4}{3\cdot2} = \frac{1}{3};$$

$$\eta_{э1(нм)} = \frac{6\cdot1-2\cdot3-4}{3\cdot2} = -\frac{2}{3}.$$

Thus, for a prediction using three blips per scan the extrapolation algorithm for the nonmaneuvering target will first have the form:

$$x_{э(нм)} = x_{3+1} = \frac{4}{3}x_1 + \frac{1}{3}x_2 - \frac{2}{3}x_3. \qquad (5.18)$$

It is easy to reckon that the weight coefficients for the maneuvering target during extrapolation in three sequential scans will be equal: $\eta_{э3(м)} = 3$; $\eta_{э2(м)} = -3$; $\eta_{э1(м)} = 1$.

The extrapolation algorithm in this case will have the form:

$$x_{э(м)} = x_{3+1} = 3x_1 - 3x_2 + x_3. \qquad (5.19)$$

Let us examine examples of use for these algorithms, assuming that three blips with coordinates $x_1 = 6$, $x_2 = 4$, $x_3 = 2$ are received from the nonmaneuvering target, and from the maneuvering target $x_1 = 9$, $x_2 = 5$, $x_3 = 2$.

Then, the extrapolated coordinate for the nonmaneuvering target, in accordance with expression (5.18), will be defined as

$$x_{э(н)} = \frac{4}{3} \cdot 6 + \frac{1}{3} \cdot 4 - \frac{2}{3} \cdot 2 = 8,$$

and the extrapolated coordinate for the maneuvering target, in accordance with expression (5.19), will be defined as

$$x_{э(м)} = 3 \cdot 9 - 3 \cdot 5 + 2 = 14.$$

Algorithms of this type are rather simply realized on computers. Single deficiency is that with an increased number of blips, by which the next coordinate is predicted, larger computer storage capacity is necessary.

In the extrapolation process errors appear and the predicted coordinates, with only some assumptions, determine the true position of the next blip. Accuracy in predicting depends on the accuracy in determining the coordinates. The greater the errors in measuring them the worse the prediction. Errors increase noticeably, when the target maneuvers.

Even the processing methods also affect the accruacy of the extrapolation. First of all this concerns the selection of the extrapolation algorithm. Thus, the algorithm, the basis of which is the hypothesis of the rectilinear and uniform movement of the target, will give the greatest errors in those sectors where the target performs a maneuver.

The magnitude of an extrapolation error largely depends on the extrapolation interval: the greater the time the target position is predicted, the more intense the increase in prediction errors. This develops especially when the target is maneuvering. The selection of the prediction interval should be reasonable.

The number of blips used for processing will also affect extrapolation errors. With an increased number of blips, the extrapolation errors are reduced, but computer storage capacity should correspondingly be expanded, which is not always possible. Figure 5.3 gives the dependences of the rms extrapolation errors upon the number of blips used for the extrapolation (a) and upon the prediction interval (b).

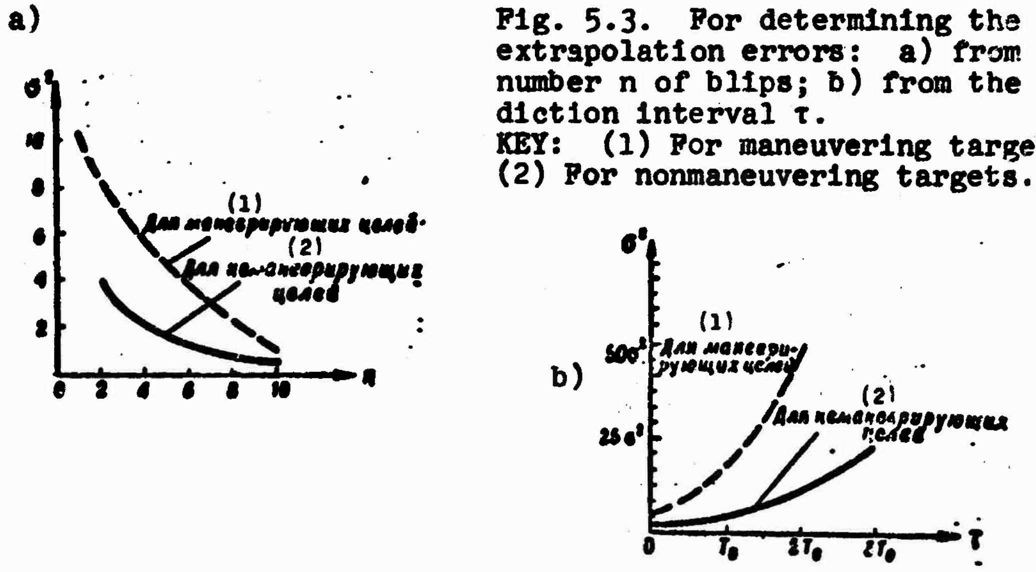


Fig. 5.3. For determining the rms extrapolation errors: a) from the number n of blips; b) from the prediction interval τ.
KEY: (1) For maneuvering targets; (2) For nonmaneuvering targets.

The particular case of extrapolation when the prediction interval is taken as equal to zero (i.e., there is no prediction) and the improved position of the blip admitted fast is called smoothing the trajectory. A decrease in errors in determining the coordinates of the current blip during smoothing is reached because of the use of the information from the previous blips. For the purpose of smoothing it is possible to use the algorithm

$$x_{\text{с}}=\sum_{i=1}^{n}\eta_{\text{с}i}x_{i},\tag{5.20}$$

which unlike the extrapolation algorithm has other coefficients. These coefficients for the case of the nonmaneuvering t--get are found from the exrpession

$$\eta_{\text{сн}i}=\frac{6i-2n-2}{n(n+1)}.\tag{5.21}$$

The smoothing algorithm for a nonmaneuvering target with three blips will have the form:

$$x_{\text{снм}}=\frac{5}{6}x_1+\frac{1}{3}x_2-\frac{1}{6}x_3.\tag{5.22}$$

During a maneuver (during a coordinate change in conformity with the polynomial of the second degree), the smoothing weight coefficients are found from the expression

$$\eta_{\text{см}i}=\frac{3}{n(n+1)(n+2)}\left[(n+1)(n+2)-2i(4n+3)+10i^2\right].\tag{5.23}$$

The smoothing algorithm for a maneuvering target with four blips will have the form:

$$x_{\text{см}}=-\frac{1}{5}x_1+\frac{1}{10}x_2+\frac{2}{5}x_3+\frac{7}{10}x_4.\tag{5.24}$$

A simplified block diagram of the resolver for realizing the obtained smoothing or extrapolation algorithms using three coordinate values is given in Fig. 5.4. The diagram includes the memory unit (delay lines $Л3_1$ and $Л3_2$) intended for storing the last two values of the coordinate, equipment for realizing the weighting function (η_1, η_2, η_3), and an accumulator (Σ).

The last value of coordinate x_1 is weighed in conformity with weight coefficient η_1 and, furthermore, enters the input of the first delay line. The values of coordinates x_2 and x_3 obtained in the second and third scans after a delay for T_0 and

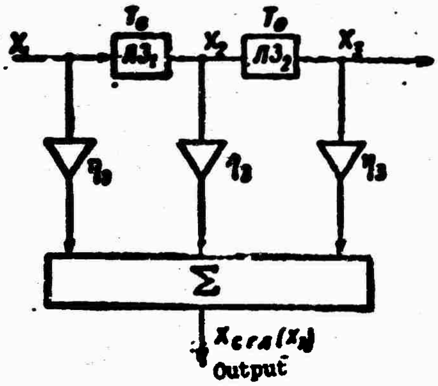


Fig. 5.4. Simplified block-diagram of the resolver for smoothing or extrapolating a coordinate.

$2T_0$ are multiplied by their weight coefficients. The obtained values simultaneously enter the input of the accumulator, at whose output the values of the smoothered or extrapolated coordinate is obtained.

The extrapolation and smoothing algorithms of the coordinate obtained above are rather complex in realization. The difficulties are basically connected with the need for having individual circuits for processing each of the independent coordinates. It is more simple to extrapolate and smooth precalculated course and speed of the target. By using the smoothed speed and course values one can clearly determine the smoothed and extrapolated coordinate values. With this method the volume of computations is considerably reduced, since instead of smoothing four quantities (x, y, v_x, v_y), only two ($V_ц$ and $Q_ц$) are smoothed. Furthermore, with this method of processing the appearance of a target maneuver is facilitated.

As is known, a target can maneuver either in terms of course or speed or simultaneously with respect to both parameters. With the independent processing of each of the coordinates it is difficult to determine the kind of maneuver, but in the case of selecting $V_ц$ and $Q_ц$ as operating parameters, the kind of maneuver is determined considerably easier. Furthermore, in a steady maneuver regime one of the parameters of the trajectory usually remains more or less constant, and the second changes according to linear law. This makes it possible during a one-parameter maneuver to continue smoothing the other parameter. During the uniform and rectilinear flight of the target, the course and speed are constant, therefore the smoothing process is very simple.

With the smoothing of the trajectory parameters, all of their values with identical weight regardless of the remoteness of admission are used. Analysis of the real trajectories of targets shows that even in sectors where there is no preconceived maneuver, the correlation of subsequent values $V_{\text{ц}}$ and $Q_{\text{ц}}$ with their previous values decreases exponentially. Thus, to smooth trajectory parameters it is advantageous to apply a method in which the previous values are considered with exponentially-decreasing weights.

In the simplest case where the course and the speed of the target are constant, to find their next value smoothing only the current measurements and the previous smoothed values are used:

$$\begin{aligned} Q_{nc} &= (1-k_c)\,Q_n + k_c\,Q_{(n-1)c}; \\ V_{nc} &= (1-k_c)\,V_n + k_c\,V_{(n-1)c}. \end{aligned} \tag{5.25}$$

where k_0 - the smoothing coefficient.

Figure 5.5 presents graphs of weight function η_1, at different values of k_c. The weight function for smoothing the course takes the form of an exponential. The less k_c, to a lesser degree are considered the results of the previous measurements (the less the smoothing), and vice versa. The exponential smoothing of the parameters can be applied in the real algorithms for smoothing and extrapolating both a nonmaneuvering and a maneuvering target.

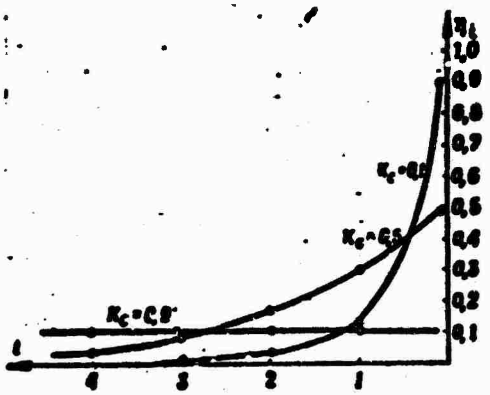


Fig. 5.5. Weight functions during exponential smoothing.

For the detection of a target maneuver the most simple to calculate are the absolute values of increments in the speed and course of the target using the formulas:

$$\begin{aligned} |\Delta V_n| &= \frac{1}{2}|V_n + \bar{V}_{n-1}|; \\ |\Delta Q_n| &= \frac{1}{2}|Q_n + \bar{Q}_{n-1}|. \end{aligned} \quad (5.26)$$

Then the squares of these increments are compared with the halves of the root-mean-square values of the random errors in the evaluation of this parameter ($0.5\delta^2_{Q_n}$ and $0.5\delta^2_{V_n}$).

If $|\Delta Q_n|^2 \geq 0.5\delta^2_{Q_n}$ or $|\Delta V_n|^2 \geq 0.5\delta^2_{V_n}$, then the decision is made that the target maneuvers in terms of course (in terms of speed), and for smoothing the course (speed) $k_{\text{CM}} = 0.1$ is taken. If these quantities are less than $0.5\delta^2_{Q_n}$, then a decision is made about the absence of a course (speed) maneuver and $k_{\text{CHM}} = 0.5$ for smoothing.

Let us examine the principle of extrapolating the coordinates using the trajectory parameters. In general form this principle can be explained thus. Let target coordinates x_n and y_n be obtained at the moment of the last scan (Fig. 5.6). Furthermore, the trajectory parameters at this point (V_n, Q_n) and their first increments (ΔV_n and ΔQ_n) are calculated. It is necessary to determine coordinates x_{n+1} and y_{n+1}.

The distance which the target will fly at time T_0,

$$l = \left(V_n + \frac{\Delta V_n}{2} T_0\right) T_0 = V_n T_0 + \frac{\Delta V_n}{2} T_0^2 \qquad (5.27)$$

The course of the target will change at this time by a magnitude of ΔQ_n and will be equal to $Q_{n+1} = Q_n + \Delta Q_n$, and speed $V_{n+1} = V_n + \Delta V_n$.

The coordinates of the extrapolated blip are:

$$\begin{aligned} x_{n+1} &= x_0 = x_n + l \sin(Q_n + \Delta Q_n); \\ y_{n+1} &= y_0 = y_n + l \cos(Q_n + \Delta Q_n). \end{aligned} \qquad (5.28)$$

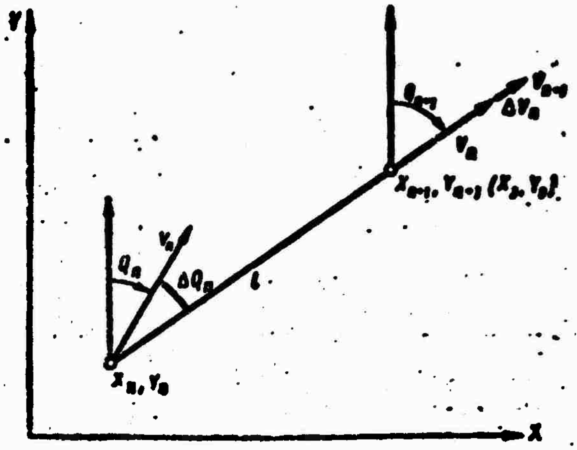


Fig. 5.6. For clarification of the principle of coordinate extrapolation with respect to the trajectory parameters.

In order to obtain parameters $Q_ч$ and $V_ч$, it is necessary to have at least two blips, but to calculate their increments ΔQ and ΔV - not less than three blips. Errors in calculating the coordinates of a blip at future positions will be determined by the errors in measuring the coordinates at the current point, as well as by the errors with which the trajectory parameters and their increments at this point are determined.

5.5. Tracking the Trajectories of Targets

Tracking the trajectories of targets consists of continuously joining the newly obtained blips to their trajectories, in smoothing the coordinates and calculating the movement parameters of the targets. If the tracking is performed automatically, it is called automatic tracking.

Let us examine the principle of realizing automatic tracking during the secondary processing of data obtained from surveillance radar. In n contiguous radar scans, blips which form the target flight trajectory are obtained (Fig. 5.7). The first problem underlying the solution during automatic tracking consists of smoothing the coordinates and calculating the parameters of the trajectories which are issued by consumers.

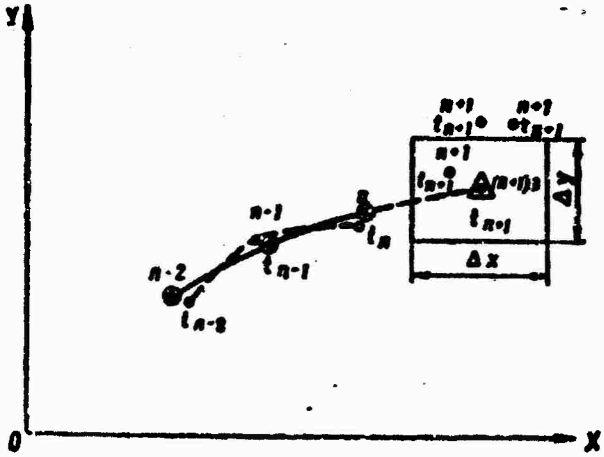


Fig. 5.7. For explantation of the means for realizing automatic target tracking.
---- - Obtained trajectory.
—— - Smoothed-out trajectory.

Let several blips be obtained in the (n + 1)-th scan. It is necessary to determine which of them belongs to this trajectory. By using the data of the known (n - 2)-th, (n - 1)-th and the n th scans and extrapolating for one scan ahead, it is possible to predict the position of the (n + 1)-th blip of the trajectory $(n + 1)_3$.

As a rule, due to instrument and calculation errors, as well as because of an expected target maneuver, the extrapolated blip coincides with none of the newly obtained blips. The relative position of the extrapolated and the newly obtained blip in the (n + 1)-th scan is random and is determined both by the accuracy of coordinate measurement and by the nature of the target maneuver. If the statistical characteristics of errors in coordinate measurement are known and the probability characteristics of the degree of the authenticity in the obtained solution are assigned, it is possible to isolate the area around the extrapolated blip (for example in the form of rectangle with sides Δx and Δy). This area is acceptably called a strobe. The coordinates of the center of the gate coincide with the coordinates of the extrapolated blip. If the dimensions of the gate are selected so that the probability of hitting its true blip is great, the blip which hit the gate should belong to this trajectory.

Evidently, the greater the dimensions of the gate, the higher the probability of hitting its true blip. However, in this case the probability of hitting the gate and false blips increases. Decreased strobe dimensions increase the selectivity of the process of blip sampling, but in this case the probability of missing the true blip increases.

In cases where not one, but several blips strike the gate, a new problem arises - the extraction of the true blip inside the strobe.

Finally, a situation can arise where there will not be a single blip in the gate. Then the extrapolated blip is taken as the true one.

Thus, in automatic tracking process the following operations are performed:

- **smoothing the coordinates and determining the parameters of the trajectory (course, speed, acceleration, etc.);**

- **the extrapolation of the target coordinates for the next scan or for several scans ahead;**

- **the extraction of the gate in which the appearance of a new blip is expected with certain probability.**

- **comparing the coordinates of the extrapolated blip with the coordinates of the blips which dropped into the gate and the selection of one of them to extend the trajectory.**

According to the nature of the performed operations, automatic tracking is reduced basically to solving the problems of parameter computation and to selecting true blip from the false one.

5.6. Gating and Selection of Blips in Gates

One of the basic operations performed during automatic tracking is the selection of blips for the extension of each of the trajectories being tracked. The problem here consists in "tying" the blips observed in the next radar scan to the trajectories being tracked. This operation is performed by comparing the target coordinates and movement parameters obtained in the next current scan with the extrapolated coordinates and characteristics of the tracked trajectories. To simplify the process of trajectory selection, the coordinates of the observed and extrapolated blips are compared only in the gates.

Gating can be physical and mathematical. Physical gating is the extraction of the probable area of appearance for a blip belonging to the tracked trajectory by means of the direct effect on the radar receiver. In this case the output of the receiver is triggered only in the area of the probable appearance of a blip. Mathematical gating is the formation of the probable area of the appearance of a blip in the form of an unknown total set of numbers (the gate boundaries).

During the processing of data in the polar coordinate system the gate is set by two range values and by two azimuth values, which determine the gate boundaries (Fig. 5.8a), or by the coordinates of the center of the gate and by its dimensions relative to the center.

During the processing of data in the rectangular coordinate system the gate is set by two pairs of coordinates, which determine the gate boundaries, or by the coordinates of the center of the gate and by its dimensions relative to the center (Fig. 5.8b).

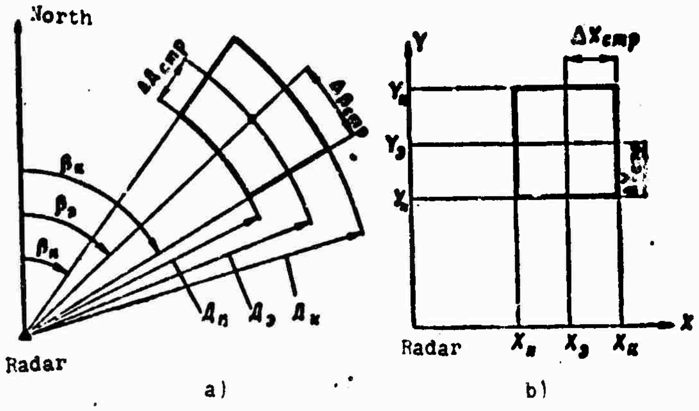


Fig. 5.8. Gates: a) in the polar coordinate system; b) in the rectangular coordinate system.

During the sorting of blips into the polar gate the following inequalities are checked:

$$\begin{gathered}|\text{Д}_i-\text{Д}_0|<\Delta\text{Д}_{\text{стр}};\\ |\beta_i-\beta_0|<\Delta\beta_{\text{стр}}\end{gathered} \tag{5.30}$$

and during the sorting of blips into the rectangular gate:

$$\begin{gathered}|x_i-x_0|<\Delta x_{\text{стр}};\\ |y_i-y_0|<\Delta y_{\text{стр}}\end{gathered} \tag{5.31}$$

The blips which satisfy inequalities (5.30) and (5.31) can be the extension of the trajectory. The separation of a singular blip from all those available in the gate is performed during the selection of blips inside of the gate.

The dimensions of the gate are chosen from the condition that an assigned probability of hitting into its true blips is provided. To obtain this probability close to unity, under the normal law of distribution of cumulative errors it is necessary to assume the gate dimensions equal to six root-mean-square cumulative errors in measuring this coordinate, for example:

$$\begin{gathered}x_{\text{в}}-x_{\text{н}}=6\delta_{x\Sigma};\\ \text{Д}_{\text{в}}-\text{Д}_{\text{н}}=6\delta_{\text{Д}\Sigma}\end{gathered} \tag{5.32}$$

The cumulative errors in measuring coordinates $\delta_{x\Sigma}$, $\delta_{y\Sigma}$, $\delta_{\text{Д}\Sigma}$, and $\delta_{\beta\Sigma}$ depend upon target maneuver, extrapolation errors, instrument errors in coordinate measurement, upon the passage of blips and upon a number of other factors. All of this leads to increased gate dimensions. Thus, in automatic tracking systems at least three gate sizes should be developed for the selection of trajectories:

1) a narrow gate for nonmaneuvering targets in the absence of blip admissions;

2) an average gate for the tracking of maneuvering targets in the absence of blip admissions;

3) a wide gate for the tracking of maneuvering targets in the presence of blip admissions.

The impingement of false blips into the gate creates in it an uncertain situation which requires further analysis. During analysis two approaches are possible.

1. With several blips in the gate, to extend trajectory for each of them. In this case it is evident that the trajectories by the false blips are soon broken, and the extension of the trajectory by the true blips will discontinue.

2. To select in the gate one blip whose probability of belonging to the trajectory being tracked is the greatest, and to discard the rest as false blips. Just such an approach is also examined below.

In general form the problem of selecting blips in the gate can be presented as a problem of checking the two rival hypotheses H_0 and H_1 for each blip which drops into the gate. Hypothesis H_0 maintains that this blip is false, while hypothesis H_1 - that this blip belongs to the tracked trajectory. If the likelihood functions of the hypotheses are known, then solution to the problem of selecting blips is reduced to checking the condition

$$\frac{L(\text{Д}_i, \beta_i/H_1)}{L(\text{Д}_i, \beta_i/H_0)} \geqslant K_i \qquad (5.33)$$

where Д_i, β_i - polar coordinates;

$i = 1, 2, \ldots, m$) - number of blips in the gate;

L - likelihood function;

K_i - solution threshold chosen from the condition of the minimum of an error of the first kind.

The likelihood functions of the hypotheses are easier to compose not in terms of blip coordinates, but in terms of their deviations from the gate center. Moreover, it is easier to allow for the basic statistical distinction in the deviations of the true and false blips, which consists of the fact that the law of probability distribution of the deviation of the true blips from center is normal, and the law of probability distribution of the deviation of the false blips from the center - uniform. The selection process is optimized by the criterion of maximum likelihood, in accordance with which that likelihood function of which is maximum, is taken as the true blip, i.e.,

$$L(\Delta \text{Д}_i, \Delta \beta_i) = \max.$$

This function is characterized by the two-dimensional normal law:

$$L(\Delta \text{Д}_i, \Delta \beta_i) = \frac{1}{2\pi \sigma_{\text{Д}\Sigma} \sigma_{\beta\Sigma} \sqrt{1 - r_i^2}} \exp\left\{ -\frac{1}{2(1 - r_i^2)} \times \right.$$
$$\left. \times \left[\frac{\Delta \text{Д}_i^2}{\sigma_{\text{Д}\Sigma}^2} - \frac{2 r_i \Delta \text{Д}_i \Delta \beta_i}{\sigma_{\text{Д}\Sigma} \sigma_{\beta\Sigma}} + \frac{\Delta \beta_i^2}{\sigma_{\beta\Sigma}^2} \right] \right\},$$

where r_1 - correlation coefficient of quantities $\Delta \text{Д}_1$ and $\Delta \beta_1$.

We will designate the expression in brackets λ^2. Then,

$$L(\Delta \text{Д}_i, \Delta \beta) = \frac{1}{2\pi \sigma_{\text{Д}\Sigma} \sigma_{\beta\Sigma} \sqrt{1 - r_i^2}} \exp\left[-\frac{1}{2(1 - r_i^2) \cdot \lambda_i^2} \right]. \qquad (5.34)$$

From the last expression it follows that to maximize the likelihood function it is necessary to minimize λ_1^2, i.e.,

$$\lambda_i^2 = \frac{\Delta \text{Д}_i^2}{\sigma_{\text{Д}\Sigma}^2} - \frac{2 r_i \Delta \text{Д}_i \Delta \beta_i}{\sigma_{\text{Д}\Sigma} \sigma_{\beta\Sigma}} + \frac{\Delta \beta_i^2}{\sigma_{\beta\Sigma}^2} = \min. \qquad (5.35)$$

Equation (5.35) is the equation of an ellipse whose center coincides with the center of the gate.

Thus, the optimal rule for selecting blips by the maximum likelihood method is reduced to choosing a blip, whose total elliptical deviation from the center of the gate will be minimal.

If we proceed to the rectangular coordinate system with the beginning of the coordinates at the gate center, equation (5.35) will be written in the form:

$$\lambda_i^2 = \frac{\Delta x_i^2}{\sigma_{x\Sigma}^2} - \frac{2r'_i \Delta x_i \Delta y_i}{\sigma_{x\Sigma}\sigma_{y\Sigma}} + \frac{\Delta y_i^2}{\sigma_{y\Sigma}^2} = \min;$$

$$\Delta x_i = \Delta \text{Д}_i;\ \sigma_{x\Sigma}^2 = \sigma_{\text{Д}\Sigma}^2;\ \Delta y_i = \text{Д}_i \Delta\beta_i;$$

$$\sigma_{y\Sigma}^2 = \text{Д}_i^2 \sigma_{\beta\Sigma}^2;\ r'_i = \frac{M(\Delta x_i \Delta y_i)}{\sigma_{x\Sigma}\sigma_{y\Sigma}}.$$

If we ignore the correlation of deviations, the selection of blips is simplified and is reduced to calculating the sum of the squares of the deviations along the coordinate axes for each blip which dropped into the gate:

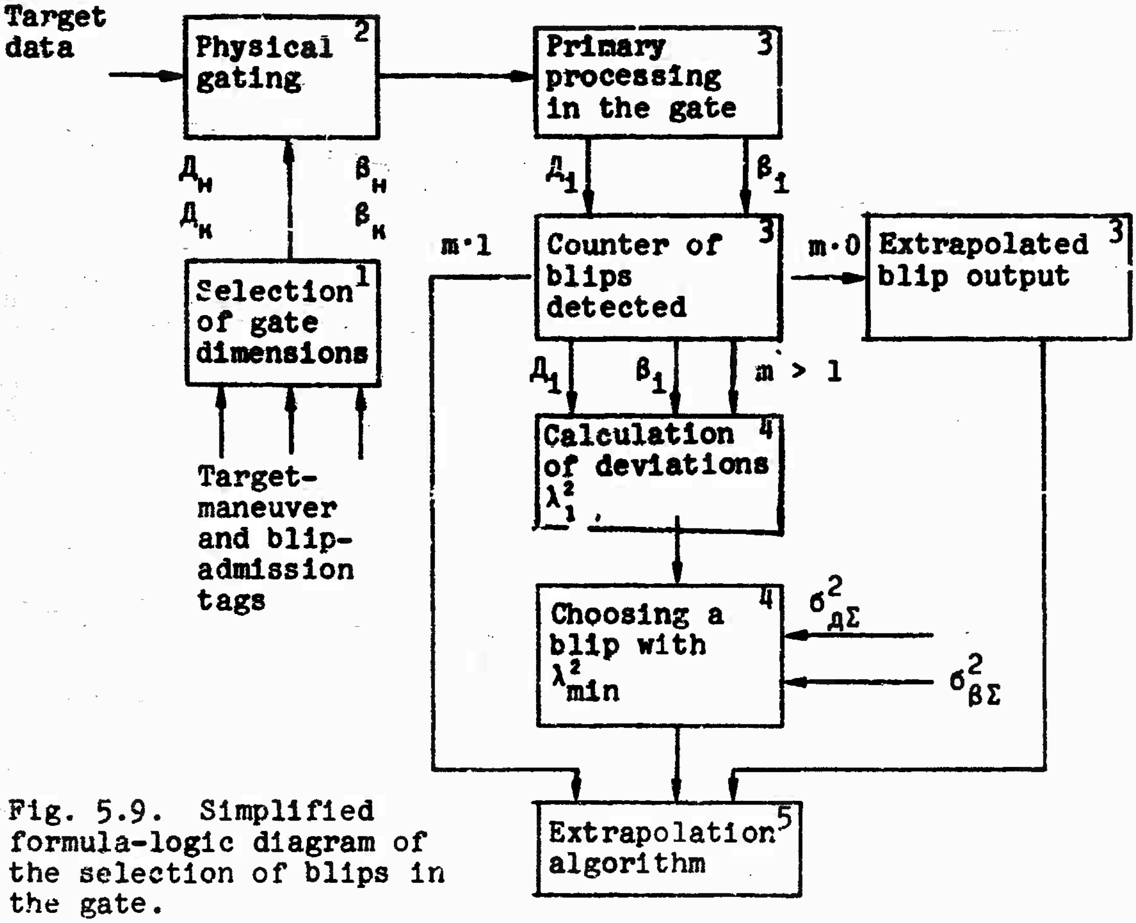
$$\lambda_i^2 = \frac{\Delta x_i^2}{\sigma_{x\Sigma}^2} + \frac{\Delta y_i^2}{\sigma_{y\Sigma}^2}, \qquad (5.36)$$

and to comparing the obtained results between them for the purpose of choosing a blip with $\lambda_i^2 = \lambda_{\min}^2$.

Furthermore, if we assume $\sigma_{x\Sigma}^2 = \sigma_{y\Sigma}^2$, then the problem is reduced to calculating the squares of the linear deviations of the blips from the gate center.

Figure 5.9 depicts a simplified formula-logic diagram of the selection of blips during physical gating in the polar coordinate system. The sequence of the algorithm fulfillment is the following.

1. The gate dimensions for the following scan (gate dimensions are chosen by taking into account target maneuver,



Target data
Physical gating 2
Primary processing in the gate 3
Д_н
Д_к
β_н
β_к
Д_1
β_1
m·1
Counter of blips detected 3
m·0
Extrapolated blip output 3
Selection of gate dimensions 1
Д_1
β_1
m > 1
Calculation of deviations λ²_1 4
Target-maneuver and blip-admission tags
Choosing a blip with λ²_min 4
σ²_ДΣ
σ²_βΣ
Extrapolation algorithm 5

Fig. 5.9. Simplified formula-logic diagram of the selection of blips in the gate.

the presence or absence of target passes in the current scan) are selected according to the processing results in the previous scan.

2. The coordinates of the gate boundaries are transmitted to the physical gating circuit of the radar receiver. The signals are removed from the receiver output only within the limits of the gate. These signals undergo primary processing.

3. The number of blips detected in the gate is computed. If there are no blips, the command is led to take the extrapolated for the true one; if there is a single blip, then it is immediately transmitted to the input of the extrapolation algorithm. With detection in the gate of several blips, the information on them

enters the calculation block to determine the normalized deviations of the blips from the center of gate.

4. The squares of the total deviations are equal and one blip with $\lambda_1^2 = \lambda_{min}^2$ is selected; the latter is led to the input of the extrapolation algorithm.

5. After extrapolating, the selection cycle is repeated.

The quality of the process of selecting blips in a two-dimensional gate can be evaluated by the probability of correct selection, i.e., by the probability of the fact that in the next scan the true blip has been selected for the extension of the trajectory. If a section is made in a rectangular gate, in which the false blips are evenly distributed with an identical average density ν, and the dimensions of the gate are chosen in such a way that the probability of the true blip hitting into the gate in the next scan equals unity ($P_0 = 1$), then the probability of a correct selection under the assumptions made is determined by the expression

$$P_c = \frac{1}{1 + 2\pi\nu\sigma_{x\Sigma}\sigma_{y\Sigma}}. \tag{5.37}$$

It follows from this expression that the probability of correct selection is greater the less the density of false blips and the less the total root-mean-square deviations of the true blip from the center of the gate.

In conclusion, we will decide on the resolving power during the selection of trajectories by means of gating. By the resolving power of the method of selecting trajectories is meant its ability not to confuse the trajectory of two close targets moving over similar trajectories. Resolving power can be evaluated by the minimum distance between trajectories at which the resolver does not entangle the trajectories with an assigned probability. The possibility of entangling neighboring trajectories is illustrated by Fig. 5.10.

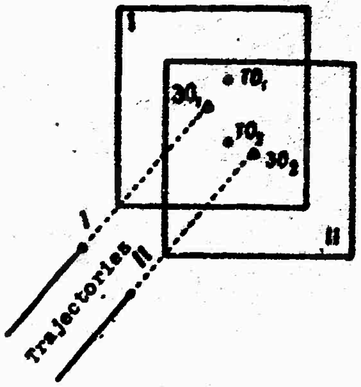


Fig. 5.10. Possible version of confusing blips during the tracking of close trajectories.

The gates for tracking trajectories I and II mutually overlap, and both blips (TO_1 and TO_2) enter the overlapped section of the gates. If the blips are sorted out by the minimum distance from the center of the gate, then they can be entangled.

The resolving power of the selection methods is usually analyzed immediately by two (even by three) coordinates. The solution of a two-dimensional (even a three-dimensional) problem is rather complex, therefore we will restrict ourselves to evaluating resolving power only by one coordinate during the sorting of blips by the linear deviation method.

If we assume that the blips enter without passes and the targets do not perform those not provided for by the algorithm for the automatic tracking of the maneuver, then the probability of entangling the trajectory is

$$P_{e} = 1 - P_{no1}P_{no2}, \tag{5.38}$$

where p_{no1} - the probability of the correct selection of the first blip for the extension of the first trajectory;

P_{no2} - the probability of the correct selection of the second blip for the extension of the second trajectory.

Since the probabilities of blip deviation from 30_1 and 30_2 are identical, $P_{no1} = P_{no2}$ and therefore

$$P_з = 1 - P_{no1}^2. \tag{5.39}$$

Using this formula a graph (Fig. 5.11) is constructed which shows the dependence of the probability of trajectory entanglement upon the normalized distance:

$$m = \frac{l_x}{\delta_{x\Sigma}}, \tag{5.40}$$

where l_x - the distance between centers of the gate at coordinate x.

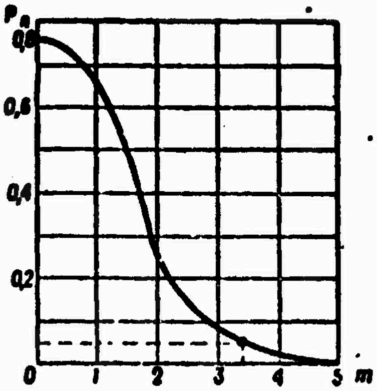


Fig. 5.11. Graph of the dependence of the probability of blip entanglement upon the normalized distance between gate centers.

For example, it follows from the graph that the trajectories are resolved with a probability of 0.95 when m = 3.5, i.e., when distance $l_x = 3.5\delta_{x\Sigma}$.

5.7. Criteria for the Quality of Secondary Processing

The information at the output of the secondary processing equipment is the initial information on the air situation and can be directly used for controlling the active means of air defense.

Naturally, the quality of troop control and the results of using the active means of air defense depend on the quality of radar information. Thus, malfunctions in tracking the target trajectories and errors in determining the coordinates and parameters of their flight lead to increased employment of air defense facilities on each target. Part of the air defense facilities can be diverted on account of the presence of false trajectories. As a result of this there are the prerequisites for the passage real targets. Thus, the quality of radar information has a whole series of requirements.

The basic criteria which characterize the quality of the secondary processing are: the accuracy in tracking the targets; the time of continuous (without malfunctions) target tracking; the number of trajectories which are being tracked simultaneously and the time of their existence; and the resolving power of the secondary processing equipment.

To quarantee high accuracy in tracking the trajectories of targets, it is necessary first to insure high accuracy in coordinate extrapolation. The accuracy of extrapolation shows the basic effect of target movement (course, speed) on the quality of parameter determination since they are produced during extrapolation. Furthermore, the effect of the process of the smoothing of the target blip coordinates which come in discretely show up on the accuracy of the issued coordinates.

It must be kept in mind that the accuracy of the output information is influenced by not only the optimality of the extrapolation and smoothing operations, but also by the optimality of the operation of comparing the blips in the gate. If the comparison is optimal, then the maximum probability of selecting the true blip from the target is guaranteed for the continuation of the route of target movement. Selecting the true blip contributes to increased output data accuracy, and, on the contrary, during selection for the continuation of the route of the false blip (or the blip belonging to another route), the output data accuracy is reduced. Moreover, under certain conditions disruption in tracking the target route during comparison can occur, when the route formed in the process of selecting false blip sharply differs from the true one, and the blips from the target cease to enter the gate. The quality of the comparison is influenced also by the size and shape of the gate.

Thus, the selection of the gate and the method of comparing the blips in the gate affect both the accuracy of the output data and the reliability of tracking the trajectories of the targets. The blip comparison method should be optimally selected on the strength of the maximum probability of selecting the true blip for the continuation of the target route. Since in this case to extend the target route either a true or a false blip is selected, it is more expedient to select an optimality criterion based on the maximum value of the likelihood ratio. Furthermore, if we fix the probability of false blip selection then we will arrive at the Neumann-Pearson criterion, which is known from the detection theory. It is completely natural therefore that the time for continuous tracking of the target trajectories depends completely on how optimal the operations of smoothing, extrapolation and comparison will be. It is measured by the magnitude of the mathematical expectation of quantity t_i - the time for the continuous tracking of the target trajectory.

The throughput of secondary processing is determined by the sequence of execution of the operations, by complexity of the processing algorithm and by the operating speed of the processing equipment. Throughput will be higher, the more simple the secondary processing algorithm and the greater the operating speed of the computer on which this algorithm is realized.

The number and duration of false target trajectories depend chiefly on the method of trajectory input for tracking. During manual input (into semiautomatic control systems), the number of false trajectories will be minimal, since the operator feeds the blip coordinates already selected by him into the machine.

During automatic lock-on the number of false trajectories is determined by the operating logic of the automatic lock-on equipment. The automatic lock-on to the target trajectories is achieved, as a rule, according to a certain number of blips following in succession, i.e., similar to the detection of signals in the logic circuits of primary processing. Hence it follows that the optimality of the lock-on process should be examined from the viewpoint of providing maximum probability in detecting actual trajectories with an assigned (permissible) probability of detecting false trajectories. Consequently, during the optimization of the lock-on process, it is advantageous to apply the Neumann-Pearson criterion.

The resolving power along the target trajectories is determined by minimum distance l between target trajectories at which the probability of "skipping" from the trajectory of one target to the trajectory of another will be greater than the assigned probability (Fig. 5.12). To ensure a high resolving power for trajectory tracking it is necessary to accurately extrapolate the target coordinates and to apply the optimal methods of comparing the blips in the gate.

Thus, the quality of secondary processing is determined basically by the accuracy of the extrapolation, smoothing and comparison operations.

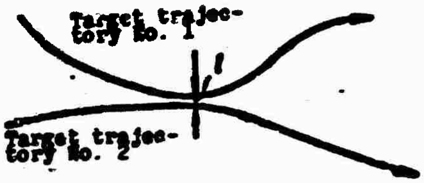


Fig. 5.12. For determining the resolving power along the trajectories.

In conclusion, let us examine the construction principle of the equipment for the automatic tracking of target routes. Tracking is performed on the basis of a computer. A simplified block-diagram of such equipment is depicted in Fig. 5.13.

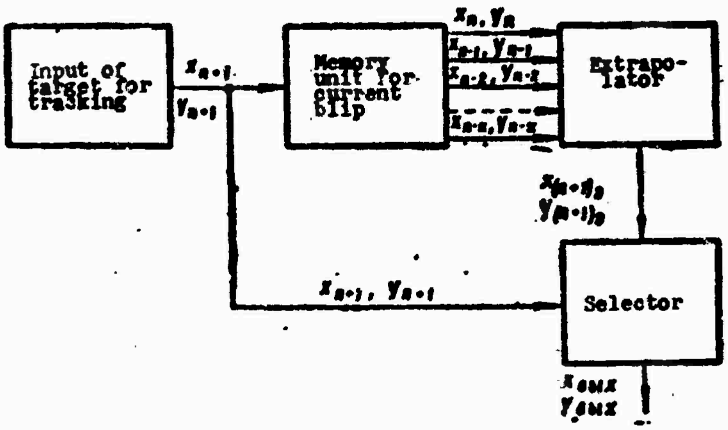


Fig. 5.13. Simplified block-diagram of the equipment for the automatic tracking of target trajectories.

The basic circuit element are the extrapolating and selecting equipment (usually realizable in the form of computer operating programs). The extrapolator calculates the target coordinates for one or several radar scanning periods ahead, using the target coordinates for several previous scans for this purpose. Data are stored in the immediate-access storage of the computer and are reviewed in each data processing cycle.

In the selector the gates are made for each individual trajectory and the situations in the gates are compared. As a result, the circuit continuously tracks the target trajectories, tying new blips to them. At the same time, the false blips are screened in the comparison process. In this sense the automatic tracking of trajectories is an additional means of screening out the false information which can enter from the output of the primary processing equipment.

5.8. Algorithm for the Secondary Processing of Radar Information

As an example, let us examine the block-diagram of the secondary processing algorithm for the American semiautomatic troop control system for the "Sage" Air Defense System (Fig. 5.14). In this case it is assumed that the information after primary processing enters the control post not only from radar, but also from other sources of information. Because of this the algorithm for the secondary processing of information repeats the individual operations of the primary processing algorithm.

The algorithm of secondary data processing begins from operator $П_1$, who inputs the data into the computer from the communication channel, which has a means of extracting the information, and control is given to operator P_2.

Logic operator P_2 determines whether a given signal is a message about a target or whether it is interference. This determination is performed by comparing the incoming message with the previously established criteria.

If the signal is a message about a target, control is given to operator P_3, but if the signal is interference, it is rejected. Logic operator P_3, by comparing the positions of a standard

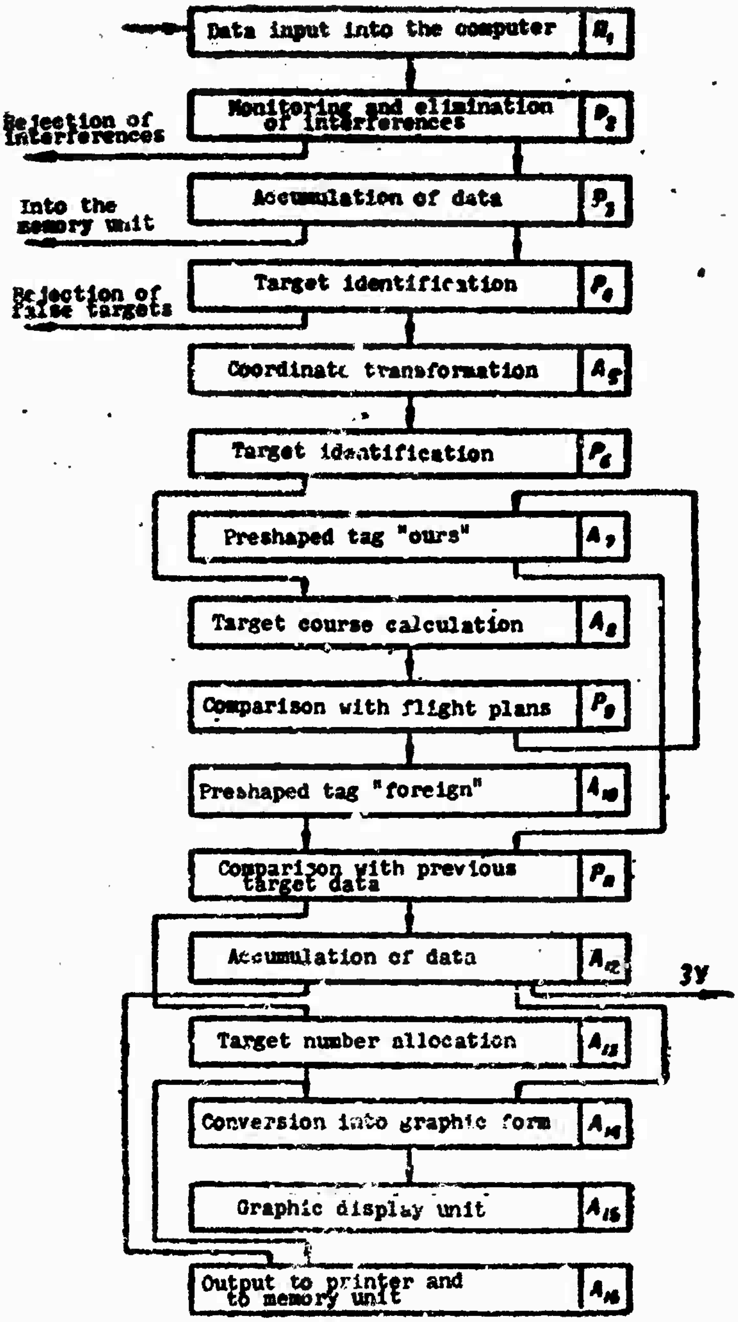

Data input into the computer
A_1
Rejection of interferences
Monitoring and elimination of interferences
P_2
Into the memory unit
Accumulation of data
P_3
Rejection of false targets
Target identification
P_4
Coordinate transformation
A_5
Target identification
P_6
Preshaped tag "ours"
A_7
Target course calculation
A_8
Comparison with flight plans
P_9
Preshaped tag "foreign"
A_{10}
Comparison with previous target data
P_{11}
Accumulation of data
A_{12}
3Y
Target number allocation
A_{13}
Conversion into graphic form
A_{14}
Graphic display unit
A_{15}
Output to printer and to memory unit
A_{16}

Fig. 5.14. Block-diagram of the algorithm for the secondary processing of information.

data card, determines whether the data is sufficient in the signal for further processing. If information is sufficient, control is given to operator P_4.

If the data are not sufficient, they are directed to the memory unit for addition with the information which entered earlier, there to be stored until the necessary minimum is accumulated information or so long as the data does not become obsolete.

Logic operator P_4 identifies the target, i.e., determines whether a given message is information about a target which interests us, and depending on the decision transmits the signal for further processing or rejects it.

Arithmetic operator A_5 converts the coordinates from the radar coordinate system, from another source of information, into a single coordinate system accepted by the control post and control is given to operator P_6.

Logic operator P_6 additionally checks the target characteristics connected with its identification, if it does not have sufficiently clear tags on its nationality. For this, it gives control to operator A_8 (target course calculation) and further to operator P_9.

Logic operator P_9 in turn compares the characteristics of the checked target with the flight characteristics of its own aircraft, which have flight planes which are in the memory unit of the computer.

If the characteristics (course, speed, altitude, time of flight, number of aircraft in the group, etc.) coincide, control is given to operator A_7, with whose help the tag "ours" is allocated to the targets, after which control is given to operator P_{11}.

If the target characteristics do not coincide with the corresponding characteristics of the targets, the data on which are contained in the flight plane, control is given to operator A_{10}, by which the targets are allocated the "foreign" tag. Thus, operator P_{11} receives the message on the target after deciding the question of who it belongs to.

Logic operator P_{11} compares the data contained in the developed message with the data that entered the control post earlier in order to establish whether data on this target was available or whether this message entered first.

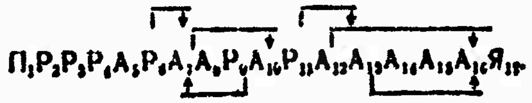
If the target data had already entered earlier, control is given to operator A_{12}, which accumulates data on the target in the memory unit, in the cell corresponding to the previously allocated target number.

From operators A_{13} and A_{12} control is given to operator A_{14}, which converts the digital message into graphic form, and then with the help of operator A_{15} this message is issued to the graphic display unit.

In parallel with the work of operators A_{14} and A_{15}, control is given to operator A_{16}, which issues the message to the printer and into the appropriate memory cell for storage.

In other automated control systems of the air defense forces, the algorithms for secondary data processing can be different, but the common form of the algorithms for computer operation in gathering and processing information will be maintained.

The operator form of writing the algorithm for secondary processing can be presented in the form



$$\Pi_1 P_2 P_3 P_4 A_5 P_6 A_7 A_8 P_9 A_{10} P_{11} A_{12} A_{13} A_{14} A_{15} A_{16} Я_{17}$$

CHAPTER 6

TERTIARY DATA PROCESSING

6.1. Definition and Makeup of Tertiary Processing

Besides the examined problems in the processing of the radar information of a single radar station, systems for controlling the combat operations of the air defense system solve an additional problem which is connected with combining the target information obtained from several radars and with creating an overall picture of the air situation.

Processing radar information that comes in from several radar stations is called tertiary data processing (TDP).

In connection with the fact that visual radar ranges frequently overlap, the information about the same target can enter simultaneously from several stations. Ideally, such target blips should be superimposed one on the other. However, in practice coincidences are not observed because of systematic and random errors in the measurement of target coordinates because of the different time of location, or because of errors which appear upon consideration of the parallel between the radar stand points and the tertiary processing point.

The chief problem in tertiary processing is resolving the question of how many targets are actually located in a controlled zone. To solve this problem it is necessary: to assemble reports incoming from the radar stations; to reduce the blips of target position to a single system of coordinates and to a single reading time; to establish the nationality of the blips to the targets; and to average the coordinates of several blips of target position to obtain one blip with more exact coordinates.

Frequently, especially in a complex air situation where there is a large number of targets in the air, the need additionally arises for the execution of one additional function included in the make-up of tertiary processing - the consolidation of information.

To solve these problems all components of the target blips are used: their coordinates, movement parameters, time of location, nationality, number of targets, number of radar stations which issued that or other radar information.

Tertiary processing equipment is comparatively simply realized by special computers which have all of the executed operations fully automated. However, sometimes to simplify automatic equipment, certain TDP operations can be performed by commands and with the participation of a human operator. In particular, to such operations can refer the identification and the averaging of blips, as well as the consolidation of information.

Tertiary processing is the concluding stage in obtaining information on the air situation.

6.2. Assembling Reports Incoming from Radar Stations

A report on targets is acceptably called information which contains information on the location of the targets, on their characteristics, etc., which is issued from radar stations through communication channels for its further processing and use.

The problem of assembling reports consists of receiving the greatest possible information with minimum losses.

To evaluate the capabilities of the equipment for report assembly, the queueing theory is used. The equipment for report assembly should consider the quenueing system.

The fundamental characteristic of such a system is the relative throughput

$$q=\frac{\bar{Q}_{обр}}{\bar{Q}_{пос}}, \qquad (6.1)$$

where $\bar{Q}_{обр}$ is the average number of processed reports; $\bar{Q}_{пос}$ is the average number of incoming reports.

The reports are usually transmitted in code. Thus, during the transmission of a report in binary-decimal code, every report consists of several digit groups (Table 6.1) to which a specific semantic value is allocated. Digit group is called a "word."

Table 6.1.

Semantic value	Service tag		Target number			Coordinates				Target characteristics		Time of location		Service tag	
Value in decimal code	9	9	5	1	2	6	8	1	1	4	0	1	6	0	0
Value in binary code	1001	1001	0101	0001	0010	0110	1000	0001	0001	0100	0000	0001	0110	0000	0000

The composition of the report includes the service tags, which determine the beginning and the end of each report. The number of digits (or bits) necessary to form a report is called the length L of the report.

If transmission time $t_{\text{пф}}$ of one digit is known, then the minimum transmission time of one report is

$$T_{\text{д}} = L t_{\text{пф}} \tag{6.2}$$

Between reports there can be pauses with average duration $T_{\text{п}}$. In some data transmission systems these pauses are constant, while in others they can be random. For such systems the average rate of report transmission is

$$V_{\text{д}} = \frac{60}{T_{\text{д}}(\text{сек}) + T_{\text{п}}(\text{сек})} \quad \frac{\text{reports}}{\text{min}} \tag{6.3}$$

Reports incoming through several channels to the tertiary data processing point form at its input a flow of reports, whose density is

$$\lambda = n\bar{V}_{\text{д}} \quad \frac{\text{reports}}{\text{min}} \tag{6.4}$$

where n - the number of data transmission channels.

The flow of reports in a system of several radar stations is unsteady, i.e., its density varies with time. Every report incoming from the input flow should be pre-processed (deciphered, corrected and recorded in the memory unit of the tertiary processing computer). The performance of these operations requires unknown time $t_{\text{зан}}$, during which the computer is engaged.

Let, at the moment the next report enters, the computer be engaged in processing the previous report. The incoming report can either leave system unprocessed or wait its turn for processing while the computer is not free or await processing for a certain, strictly limited time (for example, prior to the admission of the following report). In accordance with this, all queueing systems are broken down into three forms.

If report immediately vacates the system unprocessed, then such a system is called a failure system. If the report waits for an unlimited time to be processed, then such a system is called a waiting system. In this case the total set of reports waiting to be processed is called a queue.

If the time of retention in the queue is limited, then the system will be of the mixed type (with limited waiting). In such a system, if during waiting time $t_{\text{ОЖ}}$ the report has not been taken for processing, it receives "failure" and is lost.

It is evident that queueing systems with waiting are unsuitable for use to assemble the reports at the tertiary processing point, since with a large density of input flow the reports, which stood in queue a sufficiently long time and therefore lost their value for the recipient, will only delay the processing of later reports. It more advantageous in such a case to use the mixed type of queueing systems with a waiting time selected from the condition of the best processing of reports.

One should, however, bear in mind that when the waiting time is considerably less than the machine engagement time, the report assembly system with waiting must be considered as a failure system. An approximate block-diagram of a report assembly system with waiting is shown in Fig. 6.1.

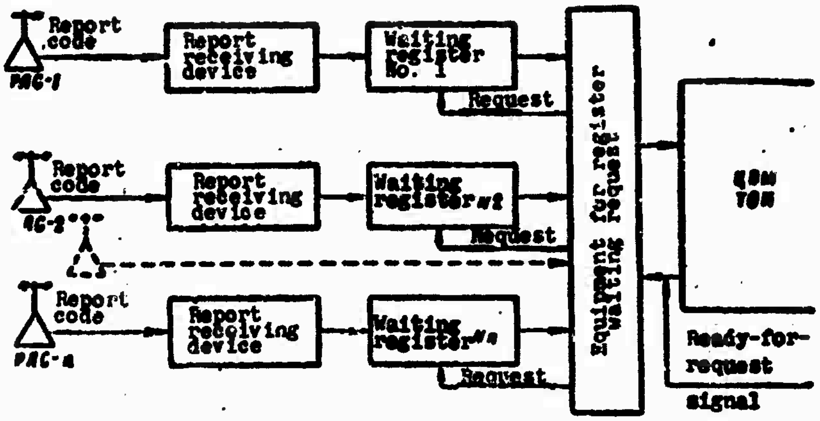


Fig. 6.1. Block-diagram of a report assembly system with waiting.

The Poisson report flux is most frequently taken as the input flux. The Poisson distribution of report flux in the queueing theory is similar to the normal distribution in the probability theory. The reasons lying as the basis of these phenomena have a similar nature.

As is known, the normal probability distribution law appears when a large number of errors approximately identical in magnitude (errors which occur for different reasons) adds up.

It has also been proven that when a large number of report flows approximately identical in density accumulates, a flow possessing the properties of the Poisson (simplest) flux is obtained.

The mathematically simplest flux can be described in different ways. The most convenient is the following definition.

The simplest flow is that flow of reports, where intervals t between any two successive reports are independent random quantities with a distribution function

$$W(t) = 1 - e^{-\lambda t}. \qquad (6.5)$$

It follows from this definition that probability P_1 of the appearance of a report in interval $(t, t + \Delta t)$ under the condition that in interval $(0, t)$ there were no reports equaling

$$P_1 = \frac{W(t+\Delta t) - W(t)}{1 - W(t)} \approx \frac{W'(t)\Delta t}{1 - W(t)} = \frac{\lambda e^{-\lambda t}}{e^{-\lambda t}}\Delta t = \lambda \Delta t. \qquad (6.6)$$

From the last expression it is apparent that the probability of the appearance of reports does not depend on time t which passed after the admission of the previous request, but is determined by interval length Δt and by density λ of the report flow.

The probability of a report not appearing in interval Δt under the same condition equal

$$P_0 = 1 - P_1 \approx 1 - \lambda \Delta t. \qquad (6.7)$$

The average time between two successive reports is

$$t_{cp} = \int_0^\infty t\,dW(t) = t\lambda e^{-\lambda t}dt = \frac{1}{\lambda}. \qquad (6.8)$$

Consequently, the average number of reports which entered per unit of time is $1/t_{cp} = \lambda$, which also made it possible to name λ the flux density of the reports. This quantity characterizes the report admission rate.

For the characteristics of the throughput of failure systems quantity q (relative throughput) is introduced. For these systems q is determined

$$q = \frac{1}{T_{зан}\lambda + 1}. \qquad (6.9)$$

Formula is valid for a case where the flux can be considered stationary and the reports enter randomly (the simplest flux), independent of one another.

6.3. Reduction of Target Location Blips to a Single System of Coordinates

When target coordinates are measured relative to the stand points the radar stations, the need arises to convert them to a single system of coordinates (usually to the coordinate system of the TDP point).

Geodetic, polar or rectangular coordinate systems can be accepted as the single coordinate system.

The most exact system is the geodetic, since in it the factor of the earth's curvature is fully considered. However, the calculations in this case are complex, and it is therefore used only when the radar stations are located at great distances from one another and from the tertiary processing point.

When these distances are small, it is advantageous to use the polar or the rectangular coordinate system with altitude correction. The calculations in these systems are rather simple, and errors resulting from the replacement of a sphere by a plane are fully acceptable for solving the whole series of practical problems.

In connection with the fact that the radar stations usually determine the coordinates in the polar system, then to convert them to a single rectangular coordinate system with the center at the processing point, the following algorithm is used:

$$\begin{aligned} X &= Д_{н}\cos\varepsilon\sin(\alpha+\beta)+x^{\bullet}; \\ Y &= Д_{н}\cos\varepsilon\cos(\alpha+\beta)+y^{\bullet}; \\ H &= Д_{н}\sin\varepsilon+\frac{Д_{н}^{2}}{2R_{э}}, \end{aligned} \qquad (6.10)$$

where X, Y, H are the target coordinates datum in the single rectangular coordinate system; x^*, y^* are the coordinates of the standpoint of the radar stations relative to the data processing point; $Д_{н}$, β, ε are the target coordinates in the polar coordinate system, which are determined by the radar stations; R_3 is the equivalent radius of the earth; and α is the convergent angle on the latitude of the processing point.

The given conversion is explained by Fig. 6.2

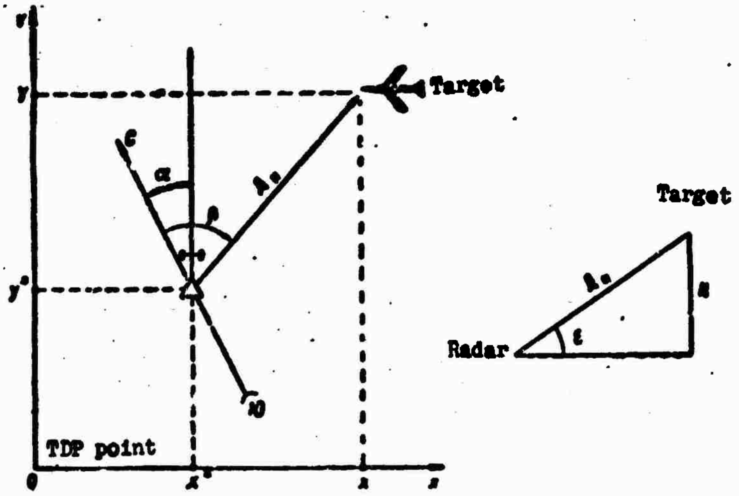


Fig. 6.2. For clarification of the conversion of the target blip coordinates from the polar system to a single rectangular coordinate system.

In automated control systems it is possible to transmit the target location coordinates from the radar to the tertiary processing point in the rectangular coordinate system. The rectangular coordinate system is also used at the processing point.

Consequently, the problem is reduced to the conversion of the rectangular coordinates of targets relative to the standpoints of the radar stations into rectangular coordinates relative to the tertiary processing point. Let us clarify the solution of this problem by the following example.

The tertiary data processing point is located at point 0, and radar - at point O_1 (Fig. 6.3). Radar determines the rectangular coordinates x_1 and y_1 of a target located at point Ц and it transmits these coordinates to the TDP point, where the coordinates are converted into rectangular coordinates relative to the TDP point (point 0). For simplicity we will consider that axes X and X', as well as axes Y and Y' are identically oriented (i.e., points 0 and O_1 are located comparatively close to one another and the convergent angle at them is identical).

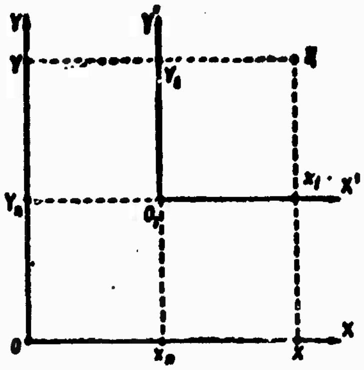


Fig. 6.3. Parallax consideration.

The conversion is performed (Fig. 6.3) using the formulas:

$$x_1 + x_p = X; \quad y_1 + y_p = Y. \tag{6.11}$$

The altitude difference of the radar and the TDP point is

$$H_{\text{РЛС}} - H_{\text{ТОИ}} = \Delta H; \tag{6.12}$$

the flight altitude of the target is

$$H = H_{\text{изм}} + \Delta H. \tag{6.13}$$

After parallax consideration it is necessary to extrapolate the coordinates, i.e., to consider the dead time of the data resulting from the removal of the coordinates in the radar, their transmission over the communication channels and the time spent on parallax consideration.

The following formulas are used to extrapolate:

$$x_{\text{э}} = X + \Delta x = X + V_x\Delta t;$$
$$y_{\text{э}} = Y + \Delta y = Y + V_y\Delta t. \qquad (6.14)$$

where V_x and V_y are components of the target speed vector (usually transmitted from radar); Δt is the time from the beginning of extrapolation to the moment the equations are solved; and x_3 and y_3 are the extrapolated target coordinates.

Now we will complicate the problem somewhat. Let us assume that from the TDP point the target coordinates proceed to the consumer in the polar system (Fig. 6.4).

Let us examine the solution of this problem only in a horizontal plane; we will convert by the successive approximations method.

From Fig. 6.4 it follows that:

$$x\cos\beta - y\sin\beta = 0;$$
$$x\sin\beta + y\cos\beta = d. \qquad (6.15)$$

Here, the unknowns are the values of horizontal range d and azimuth β, but the knowns are x and y.

The solution to the two-equation system with two unknowns by the successive approximations method is reduced to the following:

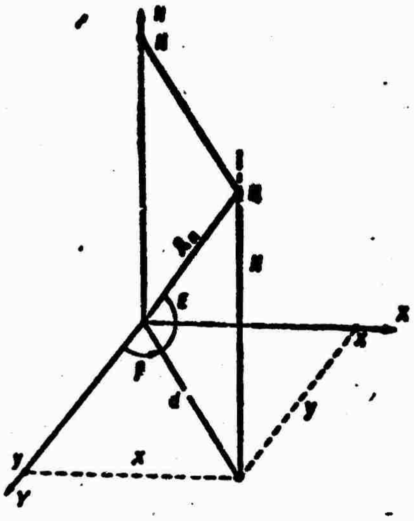


Fig. 6.4. Conversion of the rectangular coordinate system into a polar coordinate system.

- using known values x and y and unknown random value β_0 we find first approximation β_1 (i.e., instead of $\beta_{\text{ИСТ}}$ we substitute into formula (6.15) random value β_0). We will obtain

$$\begin{aligned} x\cos\beta_0 - y\sin\beta_0 &= \delta_0; \\ x\sin\beta_0 + y\cos\beta_0 &= d_0, \end{aligned} \tag{6.16}$$

where δ_0 and d_0 - linear errors in determining the azimuth and horizontal range on account of the erroneously selected β_0;

- we find the angle error (the mismatch error):

$$\Delta\beta_1 = \frac{\delta_0}{d_0} = \frac{x\cos\beta_0 - y\sin\beta_0}{x\sin\beta_0 + y\cos\beta_0}; \tag{6.17}$$

- we add the mismatch error $\Delta\beta_1$ to β_0 and we will obtain

$$\beta_1 = \beta_0 + \Delta\beta_1. \tag{6.18}$$

Value β_1 is closer to $\beta_{\text{ИСТ}}$, than β_0. This value is used as the initial quantity for the second approximation and so forth until $\Delta\beta$ equals zero (with a machine solution the approximations continue until $\Delta\beta$ becomes less than the low-order bit ($\Delta\beta_{\text{К}}$) of the azimuth code being processed in the computer).

The trigonometric functions utilized in the equations can be computed by the MacLaurin series-expansion:

$$\begin{aligned} \sin\beta &= \beta - \frac{\beta^3}{3!} + \frac{\beta^5}{5!} = \beta\left[1 - \beta^2\left(\frac{1}{3!} - \frac{\beta^2}{5!}\right)\right] = \\ &= \beta\left[1 - \beta^2\left(\frac{1}{6} - \frac{\beta^2}{120}\right)\right]; \\ \cos\beta &= 1 - \frac{\beta^2}{2!} + \frac{\beta^4}{4!} = 1 - \beta^2\left(\frac{1}{2!} - \frac{\beta^2}{4!}\right) = \\ &= 1 - \beta^2\left(\frac{1}{2} - \frac{\beta^2}{24}\right). \end{aligned} \tag{6.19}$$

On the strength of what has been said, it is possible to clearly represent the algorithm for converting the coordinates from the rectangular system to the polar system for a machine solution of the problem by the successive approximations method. The simplified block-diagram of this algorithm is presented in Fig. 6.5.

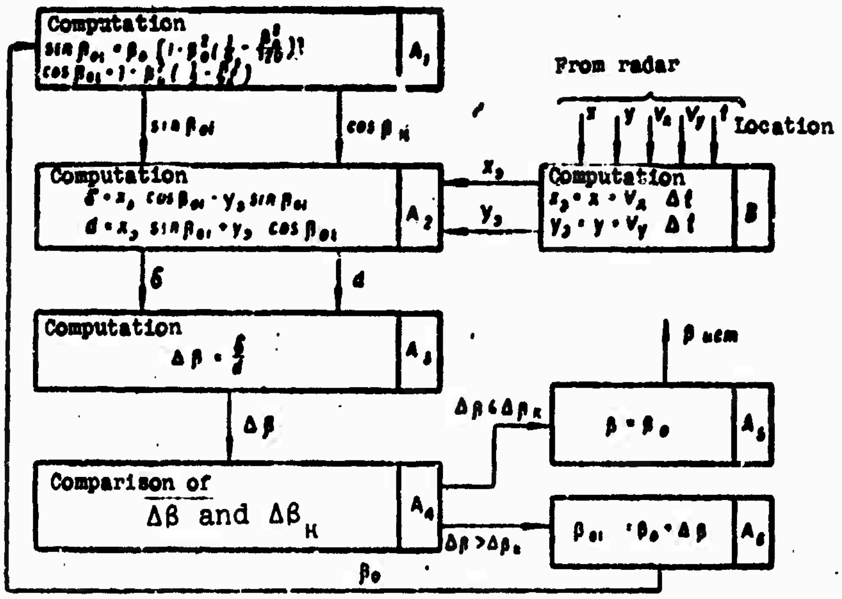

Computation
$\sin\beta_{01} = \beta_0\left[1 - \beta_0^2\left(\frac{1}{6} - \frac{\beta_0^2}{120}\right)\right]$
$\cos\beta_{01} = 1 - \beta_0^2\left(\frac{1}{2} - \frac{\beta_0^2}{24}\right)$
A_1
$\sin\beta_{01}$
$\cos\beta_{01}$
From radar
x
y
V_x
V_y
t
Location
Computation
$\delta = x_3\cos\beta_{01} - y_3\sin\beta_{01}$
$d = x_3\sin\beta_{01} + y_3\cos\beta_{01}$
A_2
x_3
y_3
Computation
$x_3 = x + V_x\,\Delta t$
$y_3 = y + V_y\,\Delta t$
B
δ
d
Computation
$\Delta\beta = \frac{\delta}{d}$
A_3
β_{ucm}
$\Delta\beta$
$\Delta\beta \leq \Delta\beta_K$
$\beta = \beta_0$
A_5
Comparison of
$\Delta\beta$ and $\Delta\beta_H$
A_4
$\Delta\beta > \Delta\beta_K$
$\beta_{01} = \beta_0 + \Delta\beta$
A_6
β_0

Fig. 6.5. Simplified block-diagram of the algorithm for converting the coordinates from the rectangular system to the polar system for a machine solution.

6.4. Reduction of the Target Location Coordinates to a Single Reference Time

The blips obtained at the TDP point from different radar stations and having different time locations are reduced to a single reference time. The single time is necessary in order to determine the position of the processed clips by condition at any one point in time. This operation considerably facilitates blip identification.

The coordinates of the identified blips are reduced to a single point in time by means of determining for every coordinate the extrapolation time relative to the assigned moment of comparison. By taking into account the comparatively high rate of information restoration, it is advantageous during extrapolation to take the hypothesis of the uniform rectilinear change in coordinates.

As an example let us examine a case where two blips M and N with time of location t_M and t_N (Fig. 6.6) have entered for processing. In order to check the nationality of the blips for one target, it is necessary to compute their position by condition for the same time. If the results of the computations prove to be sufficiently close, then it is very likely that they belong to one target.

The blips are reduced to a single time in the following manner.

Time t, for which it is necessary to reduce the blips, is assigned. The time difference for every i-th blip is found:

$$\Delta t_i = t - t_i. \qquad (6.20)$$

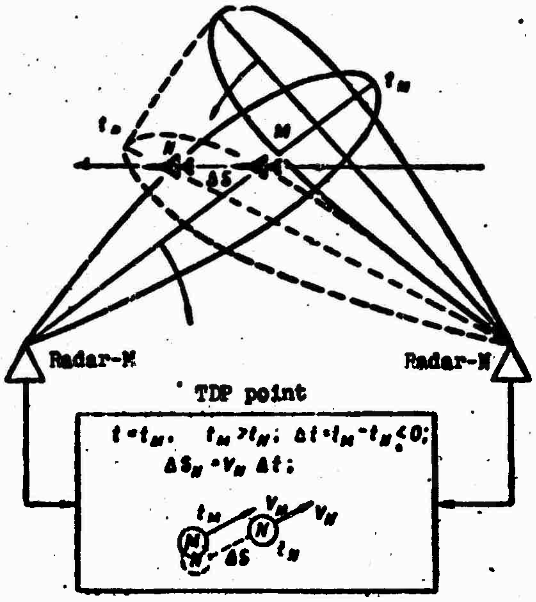

t_M
t_N
N
M
ΔS
Radar-M
Radar-N
TDP point
$t = t_M, \quad t_M > t_N; \; \Delta t = t_M - t_N < 0;$
$\Delta S_N = V_N \Delta t;$
t_M
V_M
V_N
M
N
N'
ΔS
t_N

Fig. 6.6 Explanation of the essence of reducing blips to a single reference time.

Then, the blips for period Δt_i are extrapolated, calculating route segments ΔS_i for which it is necessary to move every blip over its route:

$$\Delta S_i = V_i \Delta t_i. \tag{6.21}$$

If Δt_i is negative, then the blip moves in the direction opposite the target movement.

For the example in question time $t = t_M$ has been selected as the single reference time. Let us assume that blip N has a later time of location, i.e., $t_M < t_N$. Difference Δt in this case is negative, and blip N is displaced in a direction opposite the target movement for quantity $\Delta S = V_N \Delta t$. As a result the relative position of the blips by condition for time $t = t_M$ is obtained.

After being reduced to a single reference time and to a single coordinate system, blip identification begins.

6.5. Identification of the Blips of Target Location

All of the radar stations, regardless of each other's information, process autonomously (in their own visibility ranges) and transmit the obtained data in the form of reports to the TDP point.

Because of overlapping radar visibility ranges, in the composition of the reports there can be duplicate reports obtained from several radar stations on the same target.

In the identification process a solution is made which establishes: how many targets there actually are, if the reports about them come in from several radar stations, and how the incoming reports on the targets are distributed.

Usually the target location blips are identified into two stages: comparing the target blips - rough identification; and target blip distribution, which makes it possible by the roughly identified blips to make a more precise decision on identification.

Let us examine these stages in more detail.

The basis of the comparison stage is the assumption that reports on the same target should have identical components: coordinates, parameters, nationality, etc. By virtue of this, the decision on the nationality of the target location blips is made on the basis of comparing the components. If in this case all of the components coincide, it means the blips are identical, i.e., they belong to the same target. However, in actuality due to errors in measuring the coordinates and converting them there will be no full agreement of blips, in spite of the fact that they were obtained from the same target. As a result the uncertainty expressed by the two competing hypotheses arises.

1. Hypothesis H_1 proposes that the disagreement occurred on account of errors in the measurements and conversions, although they are from the same target.

2. Hypothesis H_2 proposes that the disagreement occurred because the blips are from different targets.

The decision to choose one or the other hypothesis is made on the basis of evaluating the magnitude of the disagreement. For this the differences in all the components comprise:

$$\begin{aligned}
x_{ij} - x_{rz} &= \Delta x; \\
y_{ij} - y_{rz} &= \Delta y; \\
H_{ij} - H_{rz} &= \Delta H; \\
V_{xij} - V_{xrz} &= \Delta V_x; \\
V_{yij} - V_{yrz} &= \Delta V_y; \\
&\cdots\cdots \\
&\cdots\cdots \text{etc.,}
\end{aligned} \qquad (6.22)$$

where i and r are numbers of radar stations from which the target location blips came; j and z are numbers of these blips; and Δx, Δy, ΔH, ΔV_x, and ΔV_y ... are obtained deviations at each of the coordinates.

Then, the permissible deviation of marks with respect to all components is established:

$$\Delta_{\text{доп}} = (\Delta X_{\text{доп}}, \Delta Y_{\text{доп}}, \Delta H_{\text{доп}}, \ldots). \qquad (6.23)$$

The limits of permissible deviations are determined by the accuracy characteristics of the coordinates from different sources, by the errors in linear extrapolation and by deviations resulting from a possible target maneuver.

The decision on target blip identification is made using the following logic. If only one of the obtained deviations exceeds the permissible value, the blips are considered to be from different targets. Otherwise they are considered identical.

The selection of the magnitudes of permissible deviations is a contradictory problem. With increased limits of permissible deviations in the blip coordinates, probability of the fact that the blips from the same target will be identified correctly increases, but at the same time the probability of their erroneous identification increases, if they are from different targets. Using known optimization methods, it is possible to determine the most advantageous magnitudes of the permissible deviations.

Quantity $\Delta_{\text{доп}}$ is calculated by the formula

$$\Delta_{\text{доп}} = K\sqrt{\sigma^2_{\text{из}} + \sigma^2_{\text{пс}}}, \qquad (6.24)$$

where K is the coefficient selected in accordance with the required probability of correct identification; when K = 3 the probability is close to 1; and $\sigma_{\text{из}}$ and $\sigma_{\text{пс}}$ are the root-mean-square errors in the measurement and conversion of the blip components.

In practice the required area of permissible deviations is usually obtained so high that even without considering target maneuver, it exceeds the resolving powers of the existing radars. Because of this, the group of identified blips, obtained in the comparison stage, can contain in its composition blips from different targets, which are close to each other. Because of this it is impossible to limit ourselves only to the comparison stage; it is necessary to perform supplemental logic processing in order to form smaller groups of identified blips belonging to a specific target. Supplemental logic processing is performed in the following stage - distribution stage.

In the distribution stage, to group the blips for individual target tags are used for their accessory to the information sources and target numbering in the system of these sources (radar).

The rules for the logical grouping of blips in accordance with the accessory of target reports to the information sources are formulated in the following manner.

1. If in the area of permissible deviations blips from the same radar station are obtained, then the number of targets is equal to the number of blips, since one station at the same moment of time cannot output several different blips from one target. Moreover it is assumed that in the group of blips being identified there are no blips from the same target, which were obtained from the same radar station in two consecutive scanning periods (period T_{06} of combining the blips is less than the scanning period of any of the radar stations linked with the tertiary processing point).

2. If in the area of permissible deviations from each radar station one blip is received, then it is considered that these blips belong to the same target. This is correct on the grounds that with the presence of a blip of the second target, only one radar station would issue the data on the two targets.

3. If an equal number of blips is received from each radar station, then, obviously, number of targets is equal to the number of blips received from one radar station, since it is highly improbable that within the limits of a small area the

radar station "would see" only its targets and "would not see" the target which the neighboring radar station observes.

4. If a different quantity of blips came in from several sources, it is assumed that the source from which the greatest quantity of blips is received gives the most probable situation. In this case the total number of targets is determined by the number of blips taken from this source.

Thus, processing the reports in the group consists of grouping the blips from several sources for one target. This problem is solved comparatively simply with the use of the first and second rules of blip grouping and considerably more difficult with the use of the third and fourth rules.

Figure 6.7 shows the enumerated versions of distributing the blips to the sources of information and the results of using the examined rules for grouping them are denoted.

Figure 6.7a shows four blips received from one source ($i = 1$; $j = 1, 2, 3, 4$). According to the first rule all these blips belong to different targets and will therefore not be subject to consolidation.

Figure 6.7b depicts three blips which belong th three different sources ($i = 1, 2, 3$; $j = 1$). According to the second rule of grouping, all these blips belong to one target. Subsequently, they should be converted into one consolidated blip.

Figure 6.7c shows four blips from two sources, two marks from each source ($i = 1, 2$; $j = 1, 2$).

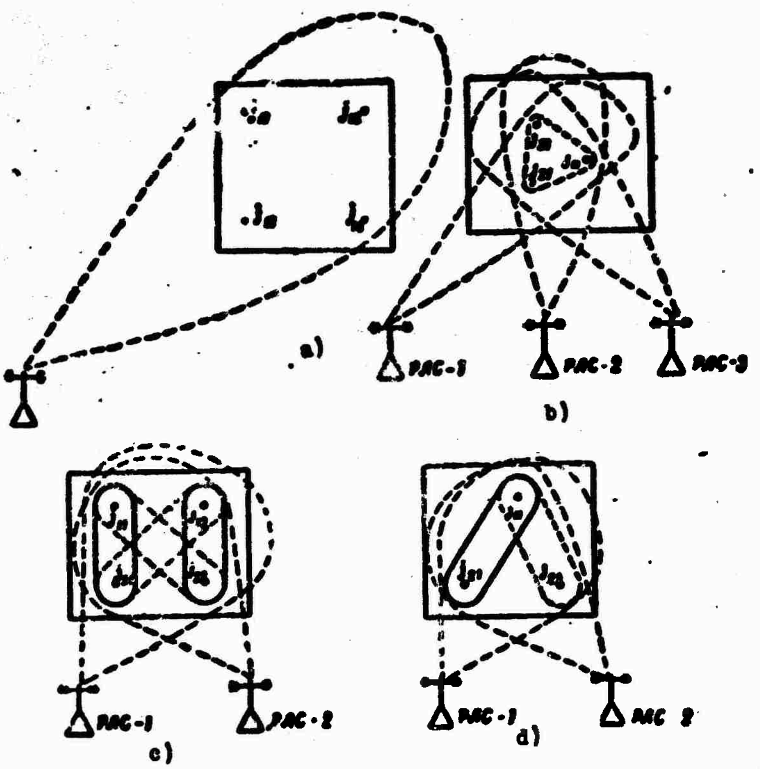


Fig. 6.7. Methods of group blips:
a - four targets; b - one target;
c, d - two targets.
Designation: РЛС = Radar station.

According to the hypothesis of the third rule we have two targets, to each of which belongs one report from each source. It is necessary to determine which pairs of blips belong to each target. The target blips can be distributed in three versions (Fig. 6.8). Version III of distribution becomes superfluous in accordance with the rule a grouping logic (reports j_{11} and j_{12} are the reports of one source, but j_{21} and j_{22} - the reports of a second source).

As for the other versions, two combinations of blip grouping are possible here:

$$\text{I}\begin{cases} j_{11} \text{ and } j_{21} \text{ - reports on the first target;} \\ j_{12} \text{ and } j_{22} \text{ - reports on the second target;} \end{cases}$$

II $\begin{cases} j_{11} \text{ and } j_{22} \text{ - reports on the first target;} \\ j_{21} \text{ and } j_{12} \text{ - reports on the second target.} \end{cases}$

The most likely version is chosen as a result of comparing the sums of the distances between blips by using the following algorithms:

$$\begin{aligned} \rho_{1121} + \rho_{1222} &= a_I; \\ \rho_{1122} + \rho_{1221} &= a_{II} \end{aligned} \qquad (6.25)$$

That combination for which the sum of the squares of the distances is minimal is accepted.

Figure 6.7d shows three blips, whereupon two of them are received from the radar station with the number i = 2, and one - from the radar station with the number i = 1.

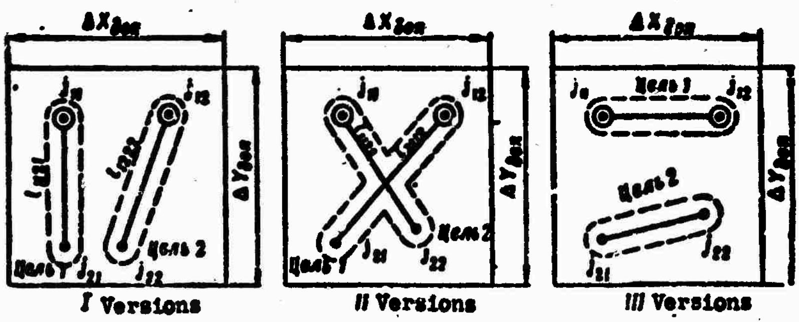


Fig. 6.8. Versions of distributing the target blips.
Designation: Цель = Target

The decision on the presence of two targets is made in accordance with the fourth rule of grouping. The possible versions of grouping blips j_{11}, j_{21}, and j_{22} for their associated target are encircled in Fig. 6.7d by continuous and broken lines.

The selection of a version is made by the same procedure as in the previous case.

The given rules of comparing and distributing the blips are unique and, depending on the required accuracy in processing, can be complicated or are simplified.

After the blips are identified, the target information is expressed by the group of blips received from several radar stations. To form one blip with more exact coordinates, the coordinates and parameters of the trajectory are averaged.

6.6. Averaging Target Location Blips

As a result of the identification, the target information represents a group of blips received from several radar stations and generally having different accuracy characteristics. For the formation from this group of one (consolidated) blip, the operation of averaging is performed.

The simplest method of averaging consists of the fact that the arithmetic mean of the coordinates is calculated using the following algorithm:

$$X=\frac{\sum_{i=1}^{m} x_i}{m}; \qquad Y=\frac{\sum_{i=1}^{m} y_i}{m}, \qquad (6.26)$$

where m - number of blips in the group.

Similar formulas are used to average the remaining parameters of the targets trajectories (altitude, speed, course angle, etc.). This method is rather simple to realize on a computer, but it does not consider the accuracy characteristics of the radar. The averaging of the target blips received from different radar stations, with consideration for the weight coefficient of the blips, is more correct. In this case this algorithm is used:

$$X=\frac{\sum_{i=1}^{m}\delta_i x_i}{\sum_{i=1}^{m}\delta_i};\qquad Y=\frac{\sum_{i=1}^{m}\delta_i y_i}{\sum_{i=1}^{m}\delta_i}, \tag{6.27}$$

where quantity δ_1, which is inversely proportional to the dispersion of errors in measuring the blip coordinates, is taken as the weight coefficient

$$\delta_i=\frac{1}{\sigma_i^2}. \tag{6.28}$$

Finally, it is possible to take the coordinates of a blip received from one radar station as being averaged, if data is available that this radar station issues accurate information. In such a case, the weight coefficient equal to unity corresponds to this source.

Besides averaging the target coordinates even the characteristics of the targets should also be generalized during the consolidation of information.

In the make-up of every report, its state accessory is transmitted as a rule, to the tertiary processing point. During the generalization of the state accessory characteristics, the machine solution is output only when correlation of the input data makes it possible to obtain reliable generalized characteristics.

If the input data are questionable (the target characteristics obtained from the individual radar stations, differ from one another), the final decision of the state accessory of the target is made by the operator.

After averaging the target blips, the obtained averaged coordinates are tied in to the consolidated trajectories of the targets, and the new averaged coordinates are allocated, the number of that consolidated trajectory, to which they belong.

CHAPTER 7

BASIC INFORMATION ON THE THEORY OF MESSAGE TRANSMISSION IN A COMMUNICATION CHANNEL

7.1. Basic Problems of the Theory of Message Transmission

The basic problem of any communication system is the effective transmission of messages from one object (or correspondent) to another. The effectiveness of transmission is understood in the sense that during the preset time through the communication system as much reliable information as possible should be transmitted. This problem can be solved with the help of systems diverse in the operating principle and design. However, as a generalized model of the single-channel automated communication system with the use of the computer it is possible to use the block diagram given in Fig. 7.1. The diagram makes it possible to indicate the general standard functions fulfilled by separate elements to establish their optimum nature.

The system of radio communication (Fig.7.2) is a special case of the presented diagram in which is understood by the message source is any object or operator producing the messages.

Messages can be divided by form into three groups: digital, continuous and mixed.

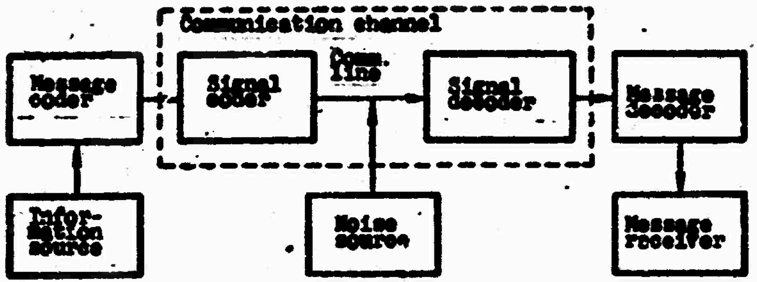


Fig. 7.1. Diagram of a single-channel automated communication system.

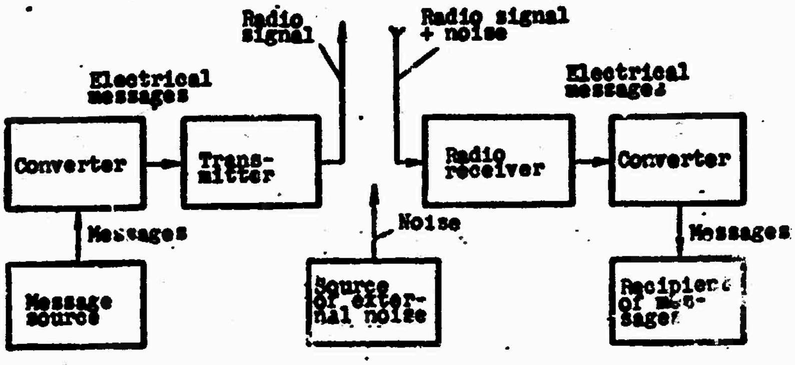


Fig. 7.2. Block diagram of radio communication.

Digital messages consist of sequences composed of a finite number of elementary characters. The complete set of elementary characters from which there is composed different messages is called an alphabet.

A typical example of the transmission of digital messages is telegraphy. Here elementary characters are pulses of two or several forms. In telemetry during the transmission of results of measurements, in the alphabet digital values of the signal (digits) can be used. In the decimal system of notation the alphabet contains ten digits (0.1, 2, ..., 9) and in the binary system - two digits (0 and 1).

Continuous messages can be changed with time, taking many values. Examples can be radiotelephony and television and also telemetry (when it is necessary to obtain results of measurements at any moment of time).

Continuous messages with a certain error can be converted into digital, for example, by means of quantization. A characteristic example is the transmission of a continuous function of time (specifically, speech) by means of pulse-code modulation.

The mixed messages consist of various combinations of digital and continuous messages.

The application of first coding equipment (the message coder is the representation of messages in a certain standard form, for example, in the form of a sequence of characters of the alphabet containing a finite number of elements. The coded messages at the exit of the coder can be presented in the form of a sequence of binary digits. The requirement that such a conversion be economical is logical, i.e., for each piece of information on the average as less binary digits as possible was used.

The second coding equipment (signal coder) serves for the conversion of the coded message into a signal being transmitted along the communication line.

During transmission various kinds of signal distortion are possible. The most characteristic distortions in radio communication are caused by the presence of disturbances, the emission of the signal through an inhomogeneous atmosphere and ionosphere (fadings) and the arrival of signals at the receiving point along several paths with a different time lag (multiwave nature).

Due to distortions the signal at the output of the communication line can be lost, and, therefore, in that case the probability of the transmission of a useful signal is indicated. The problem of the signal coder consists in the fact that in order to convert the coded messages into a signal in such a way that the errors in the determination of the correct message at the output of the communication line would be minimum.

The first decoder (signal decoder) is intended for determining the coded message by the signal taken from the output of the communication line. In the ideal case the message at the output of the signal decoder should be identical to the coded message at the input of the signal coder. If we consider this condition to be fulfilled, then the function of the second decoder (message decoder) consists in the recovery of the necessary message in its coded standard form.

The form of the message at the output of the second decoder does not have to coincide in form with the initial message of the source.

From the aforesaid it follows that the coding equipment is intended for the conversion of the message into a signal capable of being transmitted with minimum distortions through the communication line, and decoders convert the signal from the output of the communication line into the appropriate message which makes it possible to receive the correct solution.

In connection with the ordinary systems of radio communication (Fig. 7.2), the functions of two coding devices is fulfilled by the transmitter and functions of two decoders - by the receiver.

However, in the described block diagram (Fig. 7.1) two coders and two decoders are specially distinguished in order to emphasize more clearly the distinction between the two operations of coding and decoding. The advisability of such distribution can be justified by the tendency of the wide use of elements of computer technology in the contemporary systems of information transmission.

For a comparison of the communication systems with each other according to their effectiveness, it is necessary to introduce a certain measure which makes it possible to estimate quantitatively both the volume of the transmitted data (information) and the ability of the communication system to transmit these data correctly. In information theory such a universal quantitative characteristic, which does not depend on the specific physical nature of the messages, is determined by the information contained in one character relative to the other.

The value of the introduction of the concept of information is determined by two fundamental theorems of the theory of the transmission of messages [22]. The first concerns operations being fulfilled by the message coder and proves that the number of binary digits necessary on the average for the unique representation of the report is equal to the average quantity of information contained in the message.

The second theorem determines the reliability of the operations being fulfilled by the coder and decoder of the signal. It is proved that under specific conditions it is possible to code and decode signals in such a way that the probability of the erroneous reception of the transmitted message will be randomly small.

In the theory of the transmission of messages the digital and continuous communication systems are examined separately. In the digital communication system both the realization of the message and realization of the signal are sequences of characters of the alphabet containing a finite number of the latter. In the continuous communication systems realizations of the messages and signal are examined as some continuous time functions.

In this chapter the basic positions of the theory of the transmission of messages only for digital communication systems will be examined.

7.2. Concept of Information

The concept "information" in its earliest use meant knowledge by man of those or different phenomena of nature and society, i.e., it was connected with the feature of the human brain to reflect regularities of the environment. In this sense the word "information" does not lose its importance at the present time.

However, with the development of cybernetics - the science of control of complex single-minded dynamic systems - this concept obtained further development.

The concept of information in cybernetics, which widely uses the theory and technique of the transmission of messages, in a known sense is similar to the concept of repulsion in dialectric materialism.

Usually, in speaking about information, we have in mind only some properties of matter considerably important for this control process. These properties can be obtained by man, by a living organism and one or another technical equipment.

Thus, information can be considered as definite properties of matter being received by the controlling system both from the surrounding external material world and from processes occurring in the system itself. Information is necessary to provide for control, and therefore the concepts of information and control are correlative.

In the theory of the transmission of messages, by information we understand as the information about results of any event which should occur or already occurred, but its result is not previously known. The information which proceeds from the information source is included in speech, in the reading of a meter, in telephone, telegraphic and radio communications, in facsimile telegraphy, television and radar images, and so on.

As was already mentioned, the informational processes are the principle of the control. Without a knowledge of situation and without research on the information of the state one cannot proceed to the solution of any problem. Furthermore, for the solution of the problem it is necessary to know the rules of its solution, be it the algebraic equation, the firing rules or the maneuvering of forces.

As a result the processing of information of the state, according to the selected rules control instructions are produced. The control instructions are naturally also the information (instruction), which along the communication channels is transmitted from a control unit (command post, brain of man or control system of a computer) to the object to be controlled (troops, aircraft, one or another organ of man or an arithmetic unit of a computer).

As the controllable object carries out an instruction, information about this on the feedback lines enters into the control unit of the system, closing the field of the control processes. As is evident, without information there is no control.

The material form of the embodiment of the information is called a message. Messages can be represented in the form of readings of instruments, states of physical elements, printed text, digits, codes, instructions, and so on.

During the transmission of messages the random nature of their content is significant. Actually, if there would be laws which allowed the recipient to foresee precisely which message will appear at this moment, the transmission of messages would not make sense.

The message intended directly for the transmission of information over a distance is called a signal. Electrical, electromagnetic, sound, ultrasonic and other signals have obtained wide application.

7.3. Representation of Messages and Evaluation of the Quantity of Information in Equiprobable States of Elements of Messages

The large class of the digital messages being transmitted in a communication system can be presented in general form, which consists of n elements, each of which can be found in one of m different fixed states. We will call such messages digital according to the state of the elements.

Graphically the digital message can be presented in the form of the diagram depicted on Fig. 7.2, where plotted along the horizontal axis are elements of the message and along the vertical axis - the state of these elements. Let us call it the phase diagram of elements of messages or simply phase diagram.

It is important to note that the diagram reflects the most general characteristics of the transmitted report: the quantity

of elements, the total number of possible states of each element and the state of each element in this message.

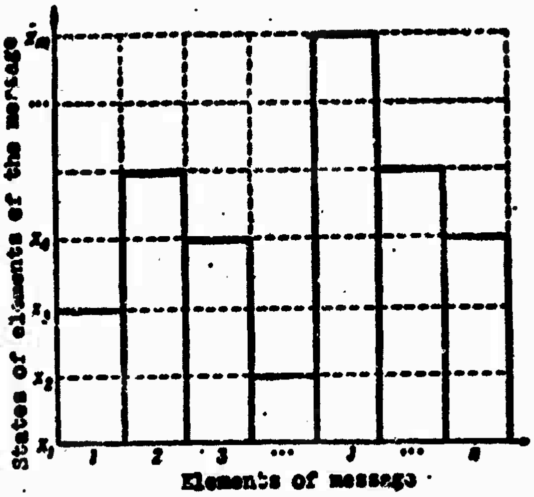


Fig. 7.3. Phase diagram of elements of messages.

Methods of the realization of messages are very diverse. Elements of the messages can be pulses of current (voltage) divided according to any criterion of selection: according to time when n pulses are transmitted consecutively along one communication channel; according to frequency when n pulses which have different carrier frequencies are transmitted in parallel along n frequency channels; and according to the position when n pulses are transmitted simultaneously along different conductors and so on.

One of the variable parameters of the pulse can be used as a state: amplitude; length; delay from the fixed point in time; polarity (for video pulses); and phase or frequency (for high-frequency pulses), and so on.

Figure 7.4 gives one message which consists of three elements (n = 3), each of which can take one of the six states.

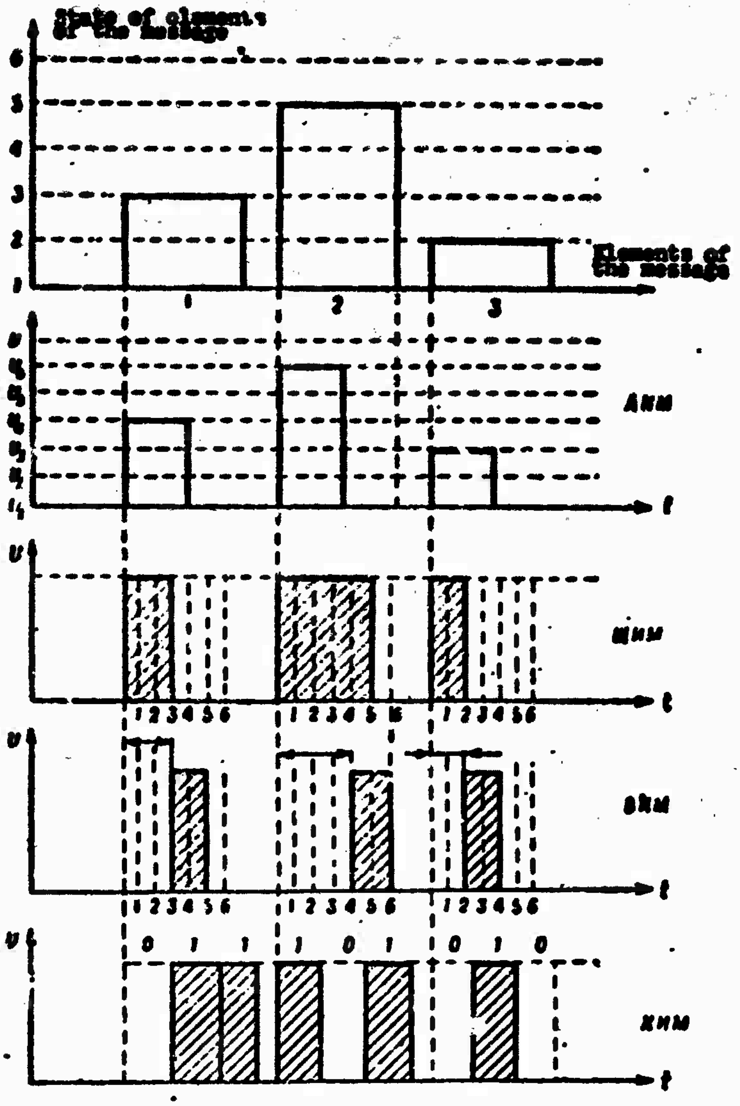


Fig. 7.4. Representation of a message by signals with different types of modulation.

Elements of the message are formed consecutively with time. The state of the elements of the message is thus determined: with pulse-amplitude modulation [AIM] (АИМ) - by the pulse amplitude; with pulse-width modulation [SHIM] (ШИМ) - by the pulse width; with time-pulse modulation [VIM] (ВИМ) - by the pulse delay with time relative to the fixed moment of time; and with pulse-code modulation [KIM] (КИМ) - by the combination of pulsed premises.

It is not difficult to visualize the elements of the messages whose states differ by modulation with respect to frequency, phase and other parameters. Let us note that sometimes the message is called a word, elements of the messages - characters, the state of elements of the message - letters, and the entire set of their possible states - an alphabet.

Therefore, the message examined above represented in general form, can be treated as a word which consists of n characters assigned by an alphabet of m letters.

Let us determine the quantity of different messages which can consist of n elements taking any of m fixed states. Let the message consist of two elements ($n = 2$), and each of the elements can be located in one of m different states.

Corresponding to the first state of the first element of the message are the states of the second element, to the second state of the first element of the message there corresponds also m states of the second, and so on. A total of m different states of the first element and m different states of the second element correspond to $L = m^2$ different messages.

If the quantity of elements of the message is equal to three ($n = 3$), then for all the different states there can be $L = m^3$, and so on.

If in a message n elements are contained, then the quantity of possible messages

$$L = m^n. \tag{7.1}$$

Quantity L can be accepted as the measure of the quantity of information. However, the selection of L as the measure of the quantity of information is connected with a number of disadvantages. For example, when L is equal to one, the message will carry zero information. This selection is inconvenient and the fact that with the addition of the quantity of information of several independent message sources, the condition of additivity, i.e., the simple linear addition of the quantity of information is not satisfied.

In fact, let us assume that the first message source is characterized by L_1 different states and the second - by L_2 different states.

Then the total number of different states for the two message sources is equal to $L = L_1 L_2$.

For the $\varkappa$ different message sources the total number of possible states is equal to

$$L = L_1 L_2 \ldots L_{\varkappa}. \tag{7.2}$$

The logarithmic measure of the quantity of information is more convenient when the quantity of information contained in the message is estimated by the logarithm of the number of possible states

$$I_0 = \log L. \tag{7.3}$$

If one state is transmitted, then the quantity of information $I_0 = 0$, which corresponds to the case when the transmission of information is absent. Having substituted the value L from (7.1) into (7.3), we obtain

$$I_0 = \log L = n \log m. \tag{7.4}$$

With an increase in the number of n elements of the message the quantity of information I_0 increases proportionally. The logarithmic measure of the quantity of information possesses the additivity in relationship to the number of elements of the message.

Let the quantity of sources of information be equal to к, then in accordance with (7.2) the total amount of information being given by all the sources is equal to

$$I_0 = \log L = \log L_1 + \log L_2 + \ldots + \log L_k \qquad (7.5)$$

or

$$I_0 = I_1 + I_2 + \ldots + I_k$$

i.e., the total amount of information is equal to the sum of the quantities of information given by separate information sources.

If $L_1 = m_1^{nI}$, $L_2 = m_2^{n2}$, ..., $L_к = m_k^{nk}$, then in accordance with formula (7.5)

$$I_0 = \log L = n1 \log m_1 + n2 \log m_2 + \ldots + nk \log m_k$$

Thus, the logarithmic measure of the quantity of information possesses additivity both in relationship to the quantity of elements in each message and in relationship to the sum of different messages.

Along with the expression for the total amount of information (7.4), it is convenient to operate by the expression for the quantity of information which is necessary for one element of the messages.

The quantity of information which is necessary for one element, the specific informativeness or the entropy of the messages

$$H = \frac{I_0}{n} = \frac{\log L}{n} = \log m.$$

In principle it is indifferent with which foundation the logarithm is selected. The different units of measure of the quantity of information will correspond to different bases of the logarithms.

If the quantity of information is calculated by means of logarithms with a base equal to ten, then the quantity of information is expressed in decimal units:

$$I_{\text{дес}} = \log_{10} L = \lg L = n \lg m.$$

By operating logarithms with base e = 2.718, we obtain the quantity of information in natural units:

$$I_{\text{нат}} = \log_e L = \ln L = n \ln m.$$

If the base of the logarithms is equal to two, then we obtain the quantity of information in binary digits or bits:

$$I_{\text{дв}} = \log_2 L = n \log_2 m.$$

It is usually accepted to select the logarithms base equal to two. In this case $H_2 = \log_2 m$, and the entropy is also measured in bits for the element of the message.

Information in the automated control systems most frequently is transmitted by binary (binary) messages, i.e., by messages of each element of which takes one of two possible messages. Each element of such a message in equiprobable states possesses one bit of information $I_0 = \log_2 2 = 1$ bit, and if the message consists of n independent binary signals with equiprobable states, then such a message contains n bits, i.e., $I_0 = n \log_2 2 =$ $= n$ bits.

Subsequently, we will use bits as the information unit, replacing the character $\log_2$ by the character log.

7.4. Determination of the Quantity of Information in the Unequally Probable States of Elements of the Messages

Basic properties of entropy. Let us examine the messages consisting of n elements, each of which is independent and can take any of the m states $X_1, X_2, \ldots, X_k, \ldots, X_m$ with probabilities $P_1, P_2, \ldots, P_k, \ldots, P_m$, respectively.

Let us assume that for a certain single message the number of elements which took states X_1 is equal to n_1, the number of elements which took state X_2 is equal to n_2, and so on. Let us note that the state $X_1 n_1$ of elements was taken with probability $P_1^{n_1}$, the state $X_2 n_2$ elements was taken with probability $P_2^{n_2}$, ..., the state $X_m n_m$ elements was taken with probability $P_m^{n_m}$. The probability of the fact that the states $X_1, X_2, X_3, \ldots, X_k, \ldots, X_m$ simultaneously take $n_1, n_2, n_3, \ldots, n_k, \ldots, n_m$ elements, according to the theorem of multiplication, is equal to

$$P = P_1^{n_1} P_2^{n_2} P_3^{n_3} \ldots P_k^{n_k} \ldots P_m^{n_m}. \tag{7.6}$$

If the number of elements making up the message is sufficiently great, then the probability of the reception n_k by elements of state X_k can be expressed in terms of the quotient equal to ratio of the number of elements which took this state to the total number of elements in the message:

$$P_k = \frac{n_k}{n}. \tag{7.7}$$

Consequently, in this case we have:

$$n_1 = nP_1;$$
$$n_2 = nP_2;$$
$$n_3 = nP_3;$$
$$\ldots\ldots;$$
$$n_k = nP_k;$$
$$\ldots\ldots;$$
$$n_m = nP_m.$$

By substituting these expressions into (7.6), we obtain

$$P = P_1^{nP_1} P_2^{nP_2} \ldots P_k^{nP_k} \ldots P_m^{nP_m}$$

or

$$P = \prod_{k=1}^{m} P_k^{nP_k}. \tag{7.8}$$

The last formula is obtained on the assumption that the number of elements in the transmitted message is sufficiently great. Actually, the existing transmitted messages do not always satisfy this requirement, and then the relation (7.7), which was used for the solution of equation (7.8), can prove to be inaccurate.

However, it is possible to examine not one separate message but a certain set of uniform messages, the total number of elements in which is sufficiently great. Then it is possible to treat the numbers n_1, n_2, ..., n_k, ..., n_m and the probabilities P_1, P_2, ..., P_k, ..., P_m corresponding to them as quantitative characteristics for the set of messages. In this case the probability P being defined by equation (7.8) can be considered as the average probability of the transmission of one message from the number of all possible messages of the set:

$$P = \prod_{k=1}^{m} P_k^{nP_k}.$$

According the average probability P of the transmission of one message it is possible to compute the average number L of all possible messages:

$$L=\frac{1}{P}=\frac{1}{\prod\limits_{k=1}^{m}P_k^{nP_k}}.$$

By knowing the average number L of all possible messages, it is not difficult to determine the average quantity of information I_0 which is contained in one message:

$$I_0=\log L=-\log\prod_{k=1}^{m}P_k^{nP_k}.$$

The obtained expression is easily converted to another form:

$$I_0=-\log\prod_{k=1}^{m}P_k^{nP_k}=-n\sum_{k=1}^{m}P_k\log P_k$$

and finally

$$I_0=-n\sum_{k=1}^{m}P_k\log P_k \qquad (7.9)$$

Having divided the average amount of information which is contained in one message by the number of its elements, we find the expression for the average entropy H of the messages:

$$H=\frac{I_0}{n}=-\sum_{k=1}^{m}P_k\log P_k=\sum_{k=1}^{m}P_k\log\frac{1}{P_k}. \qquad (7.10)$$

Quantity log $1/P_k$ can be examined as the particular entropy which characterizes the informativeness of the k-th state. At small P_k the particular entropy is great, and at large P_k (when P_k approaches one) the particular entropy tends to zero. From expression (7.10) it is apparent that the value H is the mean value of the particular entropies.

Thus, from expression (7.9) it follows that the quantity of information contained in the message depends on the number of elements of the message, the number of possible states m and probabilities of states P_k. The greatest interest is in the dependence of the quantity of information on m and P_k, since the dependence of I_0 on n is linear. Thus in more detail let us examine the main properties of entropy - the quantity of information which is necessary for one element of the messages.

First property: entropy is a value which is real, limited and nonnegative ($H \geq 0$). This property follows from expression (7.10), if we consider that probabilities P_k are nonnegative values included in the interval $0 \leq P_k \leq I$.

In spite of the minus sign, expression $-P_k \log P_k$ is a positive number. Actually, since the probability of any result $0 \leq P_k \leq 1$, then logarithm P will always be negative number ($\log P < 0$). Consequently, the product of two negative numbers gives a positive real number. Graphically the dependence of values of function $-P_k \log P_k$ on P_k is represented on Fig. 7.5.

Second property: the entropy is minimum and equal to zero if the message is known previously ($H = H_{min} = 0$). Actually, if the message is known, then it can be determined which state will take each element of the message.

Let us assume, for example, it is known previously that $P_1 = 0$; $P_2 = 0$; $P_3 = 1$; $P_4 = 0$; ...; $P_m = 0$, i.e., it is known that a certain element of the message which still is not transmitted will take the third state ($P_3 = 1$).

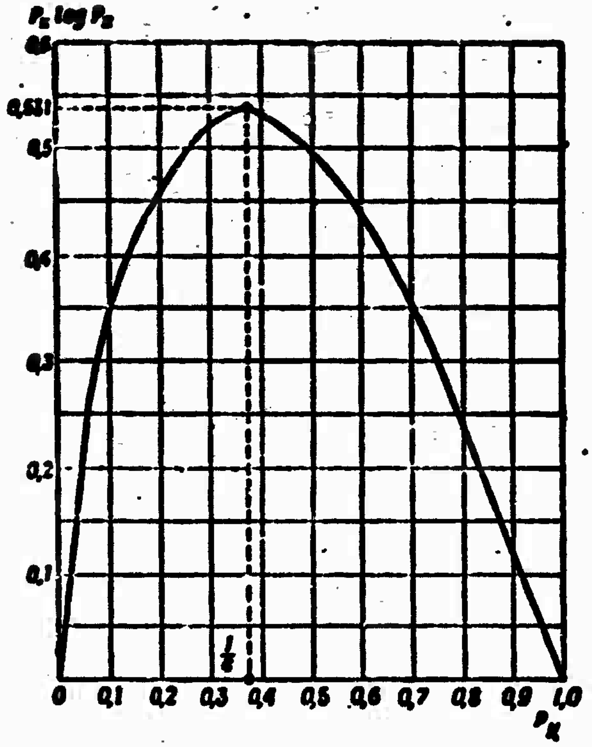


Fig. 7.5. Dependence of $-P_{K} \log P_{K}$ on the probability state P_{K}.

Thus, if during the transmission of messages there is no *uncertainty*, i.e., each message is known previously, then the entropy is equal to zero.

Third property: The entropy is maximum if all states of elements of the messages are equiprobable ($H = H_{max}$).

Let us find the value of the maximum entropy having substituted into $P_k = 1:m_2$

$$H = H_{max} = -\sum_{k=1}^{m} \frac{1}{m} \log \frac{1}{m} = \log m$$

or

$$H = H_{max} = \log m.$$

This expression coincides with that obtained earlier for the equiprobable states of elements of the messages.

Fourth property: the entropy of binary messages can change from zero to one.

The binary messages are characterized by the presence of only two possible states (m = 2). For such messages the entropy

$$H = -\sum_{i=1}^{2} P_i \log P_i = -P_1 \log P_1 - P_2 \log P_2.$$

Let us designate $P_1 = P$, then $P_2 = 1 - P_1 = 1 - P$, since $\sum_{i=1}^{2} P_i = 1$. Let us substitute the obtained values into the equation for the entropy:

$$H = -P \log P - (1 - P) \log (1 - P).$$

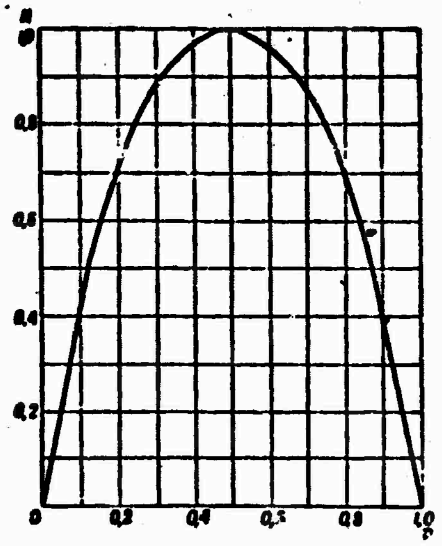


Fig. 7.6. Dependence of the entropy of a binary message on the probability of one of the states of elements of the messages.

Figure 7.6 graphically represents the dependence of the entropy of binary messages on the probability of one of the states of the elements. From the graph it is apparent that the entropy of the binary messages is changed within limits of zero to one. The entropy is equal to zero when the probability of one of the states is equal to zero, i.e., only one state is transmitted. The obtaining of only one state gives no information.

The entropy reaches a maximum equal to one if $P = 0.5$, i.e., when $P_1 = P = 0.5$ and $P_2 = 1 - P = 0.5$. In this case the uncertainty of the messages with reception is the greatest.

7.5. Coding of the Messages

The messages with the help of which the information is transmitted are usually constructed in one or another formalized language. The different elementary characters from which the messages are composed, as has already been indicated, represent the alphabet of the message. Thus, messages in the Russian language contain 33 letters of the Russian alphabet, in the Latin language - 26 letters of Latin alphabet, and so on.

In many cases it is advantageous to present the initial message of the source with the help of another alphabet. For this purpose in the diagram of the single-channel automated communication system (Fig. 7.1) there is a coder of the message whose application consists of the conversion of the input message into the sequence of characters of the alphabet selected by us.

The rule which compares each elementary character of the message to a certain combination of characters of another alphabet is called a code. The operation of the translation of the total report from one alphabet into a sequence of characters of another alphabet is called the coding of the message.

The coder converts the messages presented in the form of a certain sequence of characters of one alphabet into the combination of code characters. We will call this combination of characters of code the coded messages or simply code words.

The codes which use two elementary characters (0 and 1) are called binary codes; the codes which use three different elementary characters are called ternary codes, and so on.

At the present time in transmission information systems binary codes are widely used, since the coding and decoding equipment in this case is the simplest, cheapest and most reliable.

The same elementary message can be coded by different methods. Therefore, the question arises about the most advantageous (optimum) methods of coding. Most advantageous is such a code with which, in the first place, there is retained information contained in the message and, in the second place, on the transmission of the message the minimum of time (or, which is the same thing, the minimum number of elementary code characters of assigned length) is spent. The first requirement determines the reversibility of the operation of coding and the second - the economy of the code.

Let us explain these requirements in more detail.

It is evident that with coding, as during any other conversion, it is possible to preserve only the information contained in the message, but it is not possible to increase it. The coded message will preserve the initial information if the code words are distinguishable and uniquely connected with the appropriate initial messages.

The method of the coding with which information is not lost can be called reversible in the sense that, having used an inverse operation (decoding) to the code words, it is possible to restore the initial message without the loss of information.

Let us examine the simplest set of N mutually independent elementary messages X_i ($i = 1, 2, \ldots, N$), having a priori probabilities of transmission $P(X_i)$.

Let us assume that it is necessary to code each elementary message by means of the sequence of characters (code words) of the alphabet of the code containing L different characters v_j ($j = 1, 2, \ldots, L$), where $L < N$. When $L < N$ it is not possible to represent uniquely each elementary message by one character of the alphabet of the code, and therefore it is necessary to use several characters.

Let us note that if the alphabet contains L characters, then the number of M different sequences contained on n characters is equal to

$$M = L^n. \tag{7.11}$$

On the basis of formula (7.11) our set of different elementary messages can be uniquely coded by means of N different code words which contain n characters, where n is the smallest integer with which satisfied is the condition

$$L^n > N. \tag{7.12}$$

Different values of L and n will correspond to different codes.

Thus, for instance, for a unique representation of 32 letters of the Russian alphabet, as the binary code it is necessary to have $2^5 = 32$ different combinations. In this case all the letters can be presented in different code words composed of five binary elementary characters (0 and 1). If the elementary characters (sending of the current and pause) have an identical length, then all the code words will be identical in length. Such a code is called uniform.

The uniform code is the Baudot code used in radio telegraphy. In comparison with the nonuniform codes, the uniform codes have some advantages, in particular, they are decoded more simply, and in this case the process of decoding can easily be automated.

We illustrate the easiest method of the translation of messages (letters) into code words in the example of a binary code. This method can be reduced to the recording of different numbers in the binary system of numeration. Actually, we number all the messages according to any criterion (for example, in order of a decrease in probabilities of the appearance). Then instead of N messages it is possible to examine numbers 0, 1, 2, ..., N - 1. Let us record these numbers in the binary system of numeration.

Any number m can be recorded in the binary system of numeration in the form of a sum of the degrees of number 2:

$$m = b_k \cdot 2^k + b_{k-1} \cdot 2^{k-1} + \ldots + b_1 \cdot 2^1 + b_0$$

where b_k, b_{k-1}, ... b_1, b_0 are digits 0 to 1.

For example, $7 = 1 \cdot 2^2 + 1 \cdot 2^1 + 1 \cdot 2^0 = 111$, and $9 = 1 \cdot 2^3 + $ $+ 0 \cdot 2^2 + 0 \cdot 2^1 + 1 \cdot 2^0$.

Thus, all the messages can be numbered in the binary system of numeration, whereupon the number of the message will represent the code designation of the corresponding message.

The obtained binary code is nonuniform (code designations are formed from the different number of binary digits). With the decoding of the code words the errors and misunderstandings can appear, since in the nonuniform code we cannot understand where one code word ends and the other begins.

Thus, for instance, message No. 4 (100) can be deciphered as message No. 1 and the two following with a zero number or as message No. 2 and one message with a zero number, and so on.

For the elimination of the ambiguity between separate code words, introduced is a special separating sign which is equivalent to the transition to the ternary system of notation) or the code is made uniform, having added at the beginning of each of the code designations, length of which is less than the maximum length of the encountered code designation, the necessary number of identical digits (for example, zeros). After this a uniform code is obtained.

Let us return again to an examination of the reversible economical codes. Equation (7.12) makes it possible to assert that in principle it is always possible to indicate several codes for which the first requirement, reversibility, is fulfilled. However, as a rule, these codes will not satisfy the second requirement - the economy of the code.

The requirement for the economy of the code is very important. Radio reconnaissance can intercept information during its prolonged transmission. The transmission of the obtained information by an uneconomical code can lead to a long time delay with its delivery to the user, which is sometimes inadmissible.

The fact that many of the reversible codes will be uneconomical is proven by the following reasoning.

Let us assume that the elementary messages are letters of the Russian alphabet. Let us designate by $P(x_1)$ the probability of the frequency of the appearance of the letter X in sufficiently long semantic text. For the different letters the probabilities of their appearances are different. If these probabilities are considered, then, obviously, more economical will be that code in which the letters having great probability correspond to the shorter code designations than do comparatively improbable letters. Thus, in general the economical code is nonuniform.

In order that the nonuniform code be reversible, it is necessary that no code designation coincides with the initial part of any other longer designation. With fulfillment of this condition code designations will be different, i.e., and it is always possible to indicate where one code word ends and another begins.

For example, if in the binary code the code designation "101," is used, then no longer should there be code designations "10" or "1010." Otherwise, the beginning of the first code message can be understood as the second, the beginning of the third - as the first, and so on.

It is shown [21] that the inequality

$$\sum_{k=1}^{N} L^{-n_k} < 1$$

is the necessary and sufficient condition for the existence of N different code words using an alphabet of L different elementary characters, the length of which is equal to the assigned integers n_k, $k = 1, 2, \ldots, N$, and not one of them is an extension of a shorter word.

In information theory the following fundamental therorem about the coding of messages is proved: the minimum average number of code characters necessary for one character of the message can be made as close as desired to H: log L (i.e. to ratio of information contained in one character of message to the capacity of the alphabet of the code), applying the coding not of separate elementary messages but whole "units," which contain a sufficiently large number of elementary messages.

This theorem correlates

$$\frac{H}{\log L} < n_{cp} < \frac{H}{\log L} + \frac{1}{\nu},$$

where H is the entropy of characters of the alphabet of the messages; n_{cp} is the average of characters of the code necessary for one message; ν is the quantity of statistically independent characters which enter into one unit; and L is the quantity of characters of the alphabet.

The theorem defines the maximally possible economy of code, estimates how one or another specific code is similar to the most advantageous one, and imparts a physical sense to one of the basic concepts of the information theory (entropy of the elementary message as to the minimum number of binary characters which are necessary on the average for one such message.

From all that has been previously stated, it follows that the codes with maximum entropy, i.e., with the smallest redundancy, are the most economical from the viewpoint of transmission on communication lines. Therefore, if it is necessary to evaluate the code used from the given point of view, it is necessary to compare the entropy of this code with the maximally possible entropy for the coded alphabet (the latter will take place with equal probability of the emergence of all letters of the alphabet).

7.6. Conditional Entropy of Statistically Dependent Messages

In the solution of problems of the transmission of information, we frequently deal with several sources sending dependent messages. To evaluate the quantity of information being given by the set of such information sources, the conditional entropy has great significance.

Let us assume that there are two information sources. Elements of the first information source take states $x_1, x_2, \ldots, x_m$ with probabilities $P(x_1), P(x_2), \ldots, P(x_m)$, while elements of the second information source take states $y_1, y_2, \ldots, y_m$ with probabilities $P(y_1), P(y_2) \ldots, P(y_m)$.

The mutual statistical connection between reports X and Y can be characterized by conditional probabilities. Let us assume, for example, it is known that the first message took the state x_k. As a result of the dependence of the messages for the assigned state x_k the probabilities of the appearance of states Y are determined by the conditional probability

$$P(y_1/x_k), P(y_2/x_k), \ldots, P(y_i/x_k), \ldots, P(y_n/x_k).$$

The greater the connection between X and Y, the larger the difference will be between the largest and smallest values of the conditional probabilities. In the particular case when the statistical dependence is the greatest, the defined state x_k corresponds to one state of the set Y, for example y_i, and then one conditional probability will take the greatest value equal to one [$P(y_i/x_k) = 1$], and the remaining conditional probabilities will take the smallest value equal to zero.

Other conditional probabilities of state Y correspond to other states X. For this fixed state x_k the set of conditional probabilities determines the particular conditional entropy

$$H = (Y/x_k) = -\sum_{j=1}^{n} P(y_j/x_k) \log P(y_j/x_k), \tag{7.13}$$

which characterizes the informativeness of the messages Y since state x_k became known. In the strong statistical connection of X and Y the particular conditional entropy will be small, and, on the contrary, in the weak statistical connection - large.

If we average the particular conditional entropy (7.13) over all the states x_k, taking into account the probability of the appearance of each of the states $P(x_k)$, then we find the general conditional entropy of the messages Y relative to messages X.

$$H(Y/X)=\sum_{k=1}^{m}P(x_k)H(Y/x_k).$$

By substituting here the expression for the particular conditional entropy (7.13), we obtain

$$H(Y/X)=-\sum_{k=1}^{m}\sum_{j=1}^{n}P(x_k)P(y_j/x_k)\log P(y_j/x_k). \qquad (7.14)$$

It is known that the probability of the combined appearance of the two dependent states x_k and y_j is determined by the equality

$$P(x_k, y_j)=P(x_k)P(y_j/x_k) \qquad (7.15)$$

or

$$P(x_k, y_j)=P(y_j)P(x_k/y_j).$$

Substituting equality (7.15) into (7.14), we obtain

$$H(Y/X)=-\sum_{k=1}^{m}\sum_{j=1}^{n}P(x_k, y_j)\log P(y_j/x_k). \qquad (7.16)$$

The basic sense of the average conditional entropy (or simply conditional entropy) is the fact that it shows which entropy is given by messages Y when the entropy of messages X is already known.

Conditional entropy possesses two main properties.

First property. If messages X and Y are statistically independent, then the conditional entropy of messages Y relative to messages X is equal to the unconditional entropy of messages Y, i.e., $H(Y/X) = H(Y)$.

In this case the entire information which is contained by messages Y is new with respect to the information contained in messages X.

Second property. If messages X and Y are statistically rigidly connected, then the conditional entropy of messages Y relative to reports X is equal to zero, i.e., $H(Y/X) = 0$.

This means that the reports Y contain no new information over that which is contained in messages X.

7.7. The Transmission of Discrete Messages Along the Communication Channels with Noises and Without Noises

If in the discrete communication channel the alphabets of code characters x at the input and y at the output are identical and

$$P(y_j/x_i) = \begin{cases} 1 \text{ when } i = j; \\ 0 \text{ when } i \neq j, \end{cases}$$

i.e., the characters at the output and inlet always coincide, then such a channel is called a discrete communication channel without noise. It is completely characterized by the base of code m and by quantity of characters L being passed on the average per unit time.

The value

$$C = L \log_2 m \qquad (7.17)$$

is called the transmitting capacity of the discrete communication channel without noise.

If the information source has an entropy H (bits per character), and channel possesses transmitting capacity C (bits per second), the messages of the source can be coded in such a way as to transmit with along the communication channel with the average speed of $C/H - \epsilon$ (characters of the message per second), where $\epsilon > 0$ is as small a value as possible.

Hence it follows that in order to transmit the messages of the source along the communication channel without noise at the same speed as that which they are created by the source (i.e., without the unlimited increase in the quantity of untransmitted elements of the message), it is necessary and sufficient that the transmitting capacity C of the channel be more than the productivity of the source H'.

Actually, let us assume that the information source transmits elements of the message independently of previous ones with identical probabilities equal to 1/n, then $H = H_{max} = \log_2 n$. If $n = m$, the coding is reduced to the establishment of a one-to-one correspondence of each of the elements of the message $z_{н}$ to the character of code x_1. In this case it is possible to transmit V signs of the message per second along the communication channel.

Thus, if transmitted along the communication channel without noise is C informational units, then the maximum speed of the transmission of the message

$$V_{\max} = \frac{C}{H} = \frac{L \cdot \log_2 m}{H} \left[\frac{\text{rel. information}}{\text{s}}\right]. \qquad (7.18)$$

Practically in every communication channel as a result of different reasons there occurs noise, or, as it is called, noise which distorts the useful signals and disrupts the one-to-one correspondence of the sent and received signal. As a result it is possible to take one signal for another. The more powerful the noise, the greater the possibility for the distortion of the message.

If in the communication channel the alphabets of the code characters at the input and output are different, i.e., the characters at the output and input do not coincide, then such a communication channel is called a communication channel with noise.

Let us assume that in the communication channel (Fig. 7.1) messages with many states X and entropies H(X) are transmitted. In the communication channel messages undergo the interaction of noises, the entropy of which is equal to H(N). As a result of the interaction of noises the received messages will have Y states the entropy of which is equal to H(Y). The noises destroy the information contained in the initial messages, and at a high noise level the quantity of received information can be greatly decreased. An illustration of this can be the following example.

Let us assume that transmission will be given to binary messages possessing two states x_1, and x_2 with probabilities $P(x_1)$ and $P(x_2)$. If $P(x_1) = P(x_2) = 0.5$, then the quantity of transmitted information is maximal.

In the communication channel the noises can have such nature that the probabilities of states y_1 and y_2 at the receiving point can greatly differ, for example, $P(y_1) >> P(y_2)$, which indicates the great distortions and decrease in entropy H(Y). Distortions of such nature are observed with transmission along the communication line of signals of different polarity and noises of one polarity.

Another form of noises is possible. For example, signals can be affected by noise interference with different polarity.

In this case it can be that $P(y_1) = P(y_2)$.

However, the initial information here will be destroyed as a result of the fact that there is no one-to-one correspondence between the transmitted state x_1 and the received state y_1, and between x_2 and y_2. The statistical connection between these states will be decreased, and this connection is characterized by the conditional probabilities $P(y_1/x_1)$ and $P(y_2/x_2)$; and increasing will be the probabilities of false transitions $P(y_2/x_1)$ and $P(y_1/x_2)$, which determine the probabilities of the incorrect reception of states of the transmitted messages.

It is possible to estimate the amount of information received by the receiver in the presence of noise in the following manner. If the transmitted quantity of information is equal to $H(X)$, and the received equal to $H(Y)$, then conditions of entropy $H(X/Y)$ is that quantity of information which must be added to $H(Y)$ in order to find the entropy of the joining of X and Y. In other words, $H(X/Y)$ is that quantity of information which can give the total value of X when the quantity of information given by Y is known.

Consequently, $H(X/Y)$ can be examined as the quantity of information which is lacking for the total value of entropy of the joining with the entropy $H(Y)$ known. Therefore, $H(X/Y)$ can be called the loss of information caused by the action of noise.

If the loss of information $H(X/Y)$ is computed from the total amount of information $H(X)$, then we find the amount of information which is contained in the received set of messages Y with respect to the transmitted set of messages X. This amount of information is designated $I(Y, X)$ or $I_{Y \to X}$ or $I(Y/X)$. We take the first designation. As follows from that stated above, $I(Y, X) =$ $= H(X) - H(X/Y)$.

This expression defines the amount of information transmitted on the average along the communication channel in conditions of the action of noise. Let us examine the two limiting cases corresponding to the different effectiveness of the action of noise.

First case. The level of noise in the communication channel is insignificant or in the noise limit is completely absent. Here the messages X and Y statistically are completely dependent, and in accordance with Section 7.6 the conditional entropy $H(X/Y) = 0$.

Consequently, the amount of information is equal to the entropy of the transmitted messages:

$$I(Y, X) = H(X).$$

Second case. The noise level in the communication channel is great. At a very high noise level the messages X and Y become statistically independent. Here, as is shown in Section 7.6, the conditional entropy is equal to the unconditional: $H(X/Y) = H(X)$, and the amount of information $I(Y, X) = H(X) - H(X/Y) = 0$, i.e., the messages characterized by the Y states do not contain any information on X.

If the transmitted reports characterized by state X and the noise with states N are statistically completely independent, then the conditional entropy $H(Y/X)$ is the additional entropy which is conditioned only by the noise and is equal to $H(Y/X) = H(N)$.

Therefore, instead of $I(Y, X) = H(Y) - H(Y/X)$ we can write $I(Y, X) = H(Y) - H(N)$, i.e., for determining the amount of information contained in Y with respect to X, it is necessary to calculate the entropy of noise from the total average entropy of the received messages $H(Y)$. The indicated quantity of

information is measured in binary units per message. If we divide I(Y, X) by time T, then we determine the amount of information which is transmitted in the average message Y on messages X per unit time. This value is called the rate of information transmission along the communication channel with noise and is determined by the equality

$$R = \lim_{T\to\infty} \frac{I(Y, X)}{T} = \lim_{T\to\infty} \frac{H(Y) - H(N)}{T}.$$

The tendency T → ∞ denotes that for determining the average rate of transmission, it is necessary to examine the whole set of possible messages.

The greatest value of the transmission rate of information is called the transmitting capacity of the communication channel with noise:

$$C = R_{max} = \lim_{T\to\infty} \frac{I_{max}(Y, X)}{T}. \qquad (7.19)$$

In turn the greatest average amount of information contained in Y with respect to X,

$$I_{max}(Y, X) = \Delta F_{\text{к}} T \log\left(1 + \frac{P}{N}\right),$$

where $\Delta F_{\text{к}}$ is the transmission band of the communication channel; T is the duration of the individual message; P is the average value of power of transmitted messages; and N is the the average value of power of noise.

This amount of information can be represented as a volume in space of three measurements (Fig. 7.7). To increase $I_{max}(Y, X)$, it is necessary to increase $\Delta F_{\text{к}}$, T and P/N. Let us

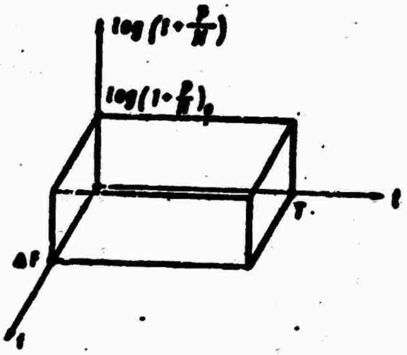


Fig. 7.7. Volume of the maximum quantity of information contained in the received messages relative to the transmitted messages.

note that the quantity $\Delta F_{\text{к}} T \log (P/N)$ is sometimes called the volume of the signal. It is obvious that the same quantity of information can be transmitted by maintaining a constant volume of the quantity of information but using different values of $\Delta F_{\text{к}}$, T and P/N.

The transmission capacity of the communication with noise ($C_{\text{п}}$) is determined in accordance with (7.20):

$$C_{\text{п}} = R_{max} = \lim_{T \to \infty} \frac{I_{max}(Y, X)}{T} = \Delta F_{\text{к}} \log \left(1 + \frac{P}{N}\right) \left[\frac{\text{binary units}}{\text{s}}\right].$$

The equation shows that the greatest rate of information transmisssion is directly proportional to the frequency ban and the logarithm of the sum (1 + P/N), i.e., it is limited by the power of the transmitter and noise. The expression obtained is called the Shannon equation and is correct for any communication system when fluctuating noise is present.

By comparing the transmission capacity in the communication channel with the noise $C_{\text{п}}$ and the transmission capacity in the communication line without noise C with an identical quantity of elementary signals (elements) of the code and equal to the speed of transmission of the elementary signals, it is visible that $C_{\text{п}} \leq C$.

Thus, a decrease in the degree of distortion of the transmitted messages and an increase in the probability of the reception of undistorted information are very important.

If there is no need to conduct transmission with high speed, then it is possible to raise the probability of reception of the undistorted information by means of its frequent repetition. In this case with an increase in the number of repetitions, the probability of the undistorted reception of the message increases. On the other hand, the higher the speed of transmission of the messages being coded by the same code, the less this probability becomes.

In practice it is very important to conduct reliable transmissions with high speed under conditions of noise. The maximum speed of the accurate transmission of information in the communication channels with noises

$$V_{\text{п}} = \frac{C_{\text{п}}}{H},$$

where $V_{\text{п}}$ is the speed of the transmission of information in the communication channel with noises; $C_{\text{п}}$ is the transmission capacity of the communication channel with noises; and H is the entropy of the transmitted message.

Just as in the case of the transmission of messages in the absence of noises, for any communication channel with noises it is always possible to select the appropriate code of the transmission of messages with close as possible to $V_{\max}$ (7.18) and with as small a probability of errors as possible in determining the transmitted signals.

If the velocity of transmission (7.19) is exceeded, then the probability of distortions will increase, and they cannot be compensated for by selection of the code.

Thus, if we select the best code, then in spite of the noises, it is possible to transmit along the communication channel the maximum quantity of information equal to the capacity of the channel, and in this case the probability of the distortions will be as small as possible. It should be noted that in this case the transmission time will increase.

It is considered that the contemporary codes are very distant from optimum, but, applying the methods of information theory, it is possible to attain their considerable improvement.

7.8. Formation of Binary Signals in the Automated Control Systems

The communication channels along which binary signals are transmitted are called binary communication channels.

Binary signals were transmitted earlier, for example, in printing telegraphy. However, the transmission of digital information, or, as one frequently says, data transmission, differs significantly from the transmission of telegraph messages. These distinctions are included in different requirements for the authenticity and speed of transmission of information.

If for telegraphy the different operation is characterized by the probability of the incorrect reception of individual characters of the order of $3 \cdot 10^{-5}$, then during the transmission of digital information frequently appear the requirements for transmission with probabilities of errors in tens and hundreds of thousands of times less than the indicated value.

The fact is that telegraphy, just as telephony, deals with the transmission of vocal messages. These messages possess great internal redundancy, which allows the recipient to restore a considerable portion of the distortions "according to sense."

(It is known that for the basic European languages the redundancy is 60-70%).

The digital information, if special measures are not provided, does not possess a redundancy which allows for the recipient to correct distortions. Therefore, even the comparatively few errors which appear in the communication channels can completely distort the content of information being transmitted to a machine or the result of its operation.

The speed of the transmission used in telegraphy also in the majority of cases does not satisfy the requirements appearing with the transmission of digital information, since the effective use of the computer in systems of automatic information processing is connected with the input and output of large volumes of information during short time intervals.

All this led to the appearance and rapid development of the trend of communications which is occupied by the development of methods and equipment of the transmission of digital information satisfying the high requirements of authenticity and speed of transmission.

Binary signals can be classed according to forms of the utilized modulation: 1) signals with amplitude modulation; 2) signals with phase modulation; 3) signals with relative phase modulation; 4) frequency-modulated signals; and 5) broadband signals which use complex forms of modulation.

The basic forms of binary signals are depicted on Fig. 7.8.

Let us examine the basic variants of binary signals and some features of their reception.

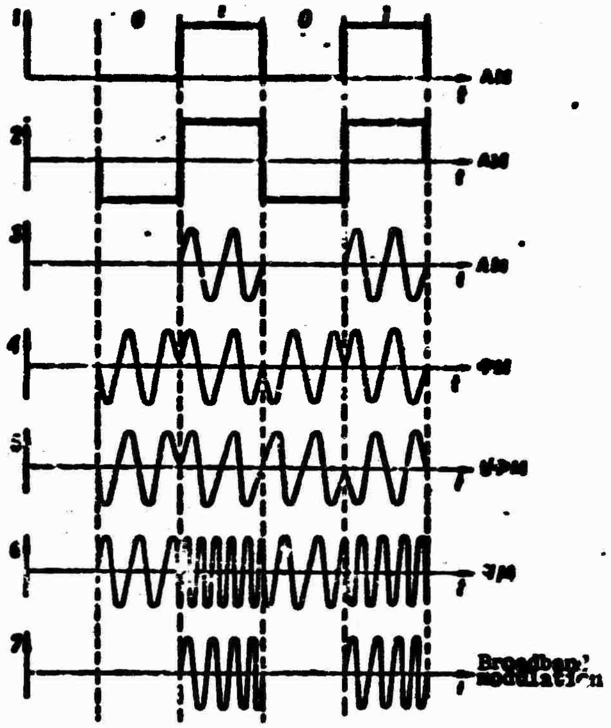


Fig. 7.8. Forms of binary signals with the different methods of modulation.

Binary signals with amplitude modulation [AM] (AM) can be unipolar (diagram 1, Fig. 7.8), ambipolar (diagram 2), or harmonic (diagram 3).

Unipolar and ambipolar signals are used for transmission along wire lines of communication and harmonics for transmission along telephone lines of communication (with a carrier frequency of 1200-1600 Hz) and along radio channels (with a carrier frequency of the radio transmitting station).

For the division of signals which correspond to zeros and ones at the output of the receiver there should be used a threshold device which issues a single signal if the voltage exceeds the threshold value and gives a zero signal if the voltage is less than the threshold.

Binary signals with phase modulation [FM] (ФМ). Typical are the signals depicted on diagram 4 (Fig. 7.8). The pulses which correspond to zero and one differ from one another only by the phase of the oscillations.

The phase is read off relative to the phase of the reference coherent voltage. If the voltage pulse which corresponds to zero has a phase different from the coherent voltage by 180°, then the voltage pulse corresponding to one has a phase coinciding with the phase of the coherent voltage. At the receiving point there occurs the conversion of the oscillations modulated in phase into oscillations modulated in amplitude.

Receiver should fulfill here two basic problems: receive the informational signals modulated in phase and form the coherent voltage.

Binary signals with relative phase modulation [OFM] (ОФМ). With phase modulation the signals are modulated in phase relative to the reference coherent voltage. With relative phase modulation the phase of each subsequent send operation is determined relative to the phase of the previous send operation.

If the subsequent send operation has a phase coinciding with the phase of the previous send operation, then 0 is assigned to the subsequent send operation. If the phase of the subsequent send operation relative to the phase of the previous send operation changes by 180°, then 1 is assigned to the subsequent send operation (Fig. 7.8, diagram 5). A reverse approach to the transmission of 0 and 1 is possible naturally.

With the OFM in the receiver two adjacent send operations are compared, but the absolute value of the phase of each, as with the ordinary phase modulation (FM) is not defined.

Binary frequency-modulated signals [ChM] (ЧМ) obtain at the present time much used and 0 and 1 are transmitted by pulses from two different frequencies (Fig. 7.8, diagram 6).

For the separate reception of the pulses, it is necessary to have two filters tuned to frequencies f_1 and f_2. At the output of the filters there is included a balance detector, which issues video signals of negative or positive polarity depending on whether the signal is received at frequency f_1 or frequency f_2 (Fig. 7.9).

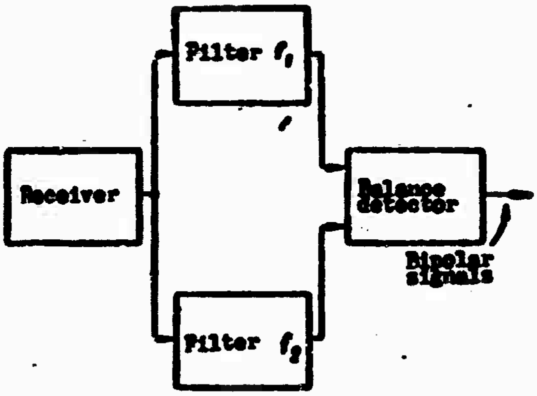


Fig. 7.9. The reception of two-frequency binary signals.

Broadband binary signals. Diagram (7) of such signals is shown on Fig. 7.8. The transmission of the zero corresponds to the absence of the signal and one - to the transmission of the pulse signal with the linearly changing frequency. Such a signal possesses a broad band of frequencies. If in the frequency band of the signal there is a narrow-band noise, then its role is less when the spectrum band of the signal is wider, since after the optimum processing of the signals the signal-to-noise ratio is proportional to ratio of energy of the signal to the spectral density of the power of the noise in the signal band.

After the optimum processing the broadband signal can be decreased in length which is considerably important for the solution of a number of problems.

7.9. Features of the Construction of Discrete Communication Systems which use Uniform Codes

In contemporary discrete communication systems, irrespective of the form of the transmitted information, as a rule, uniform codes are used. The use of uniform codes in many respects determines the features of the construction of the transmitters and receivers.

Thus, the transmitter of the discrete system, apart from coding and keying should provide, in the first place, the identical period of each transmitted pulse and, in the second place, the definite sequence of the transmission of pulses making up the code combinations. This is attained by the introduction into the transmitting unit of a system of a special distributor, which operates at a constant rate. In actuality the transmitter of the discrete communication system, as a rule, consists of a coding device, a transmitting storage element, distributor and modulator (Fig. 7.10).

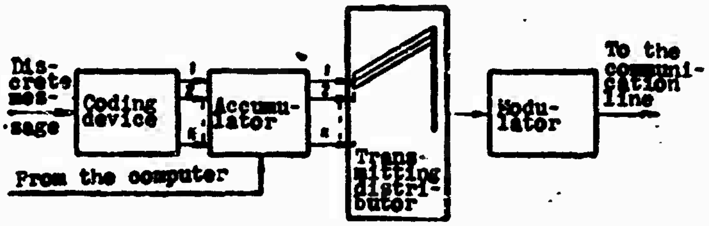


Fig. 7.10. Block diagram of the transmitter of a discrete communication system.

By means of the coding equipment the discrete information undergoing transmission is converted into a k-discharge binary number which is fixed by elements of the storage element.

The transmitting distributor, the number of contacts of which is equal to the number of elements of the code combination, consecutively converts each binary figure (0 or 1) into the appropriate electrical pulse of a definite period. Electrical signals, according to the selected method of keying are formed by the modulator.

When equipment of discrete communication is used for the transmission of results of the operation of the computer entering in binary form, the data are directly introduced into the storage element, passing by the coding device. Thus, the code combination of the transmitted message which enters into the storage element in the form of a parallel code, i.e., in the form of a code combination all elements of which are recorded simultaneously, is converted by the distributor in a sequence of pulses of equal length.

Let us now examine the conversion by the receiver of the signal into a message. From the output of the communication channel the signal is fed to the demodulator, which restores the sequence of pulses formed by the transmitting distributor.

Since the restoration on reception of the transmitted message is possible only after all k pulses are known, one of the units of the receiver of the discrete communication system should be a device which records (stores) the elements which comprise the code combinations. Such a device is called a stack and is found in any receiver of the discrete system.

It is evident that the stack should contain so many storage elements whatever the word length of the code used. In this case it is necessary that each i pulse of the received combinations would always be recorded completely by a definite i-m element of the setting device.

The selective registration of the received pulses by the elements of the setting device is provided by the receiving distributor, by means of which every element of the stack is connected in turn to the output of the demodulator.

Thus, with the help of the demodulator, distributor and stack, which are component parts of the receiver of the discrete communication system, the transmitted code combination is distinguished from the received signal and is recorded by elements of the stack.

If the received message is the information undergoing further processing in the computer, the recorded code combination at the end of each cycle of the receiving distributor is read from the stack and with a parallel code is fed to the input of the computer.

If the received messages are signs of the text, then the recorded code combinations from the stack are fed to the decoding device. The decoder provides the qualitative distribution of code combinations entering its input along the appropriate outputs.

In discrete systems with the printing of the received signs on tape or a sheet of paper, the decoding equipment, in dividing the code combinations, prepares for printing that sign, whose code combination is recorded by the stack.

Consequently, the decoder and printer together with the demodulator, receiving distributor and stack, are the main units of the discrete communication systems intended for the transmission (reception) of the text.

Figure 7.11 gives a block diagram of a receiver of the discrete communication system.

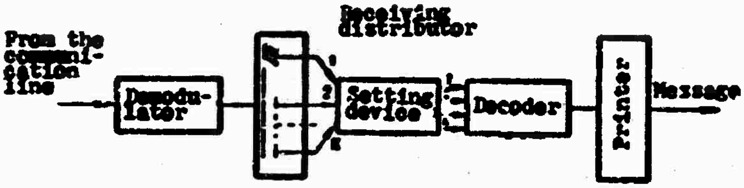


Fig. 7.11. Block diagram of the receiver of discrete communication system.

The cophasality of distributors of discrete communication systems is the basic condition of their correct operation. The loss of cophasality leads to the full curtailment of the communication. Let us explain the aforesaid in an example. Let us assume that in the time interval t_3-t_4 from the output of the demodulator the third code pulse entered. Since the third pulse should be recorded by the third storage element, then during time from t_3 to t_4 the phase of the receiving distributor should be such that connected to the output of the demodulator is the third storage element. Then during the reception of the fourth pulse (t_4-t_5) to the output of the demodulator the fourth storage element should be connected and so on. The indicated phase relationship should be retained with the reception of any code combination.

To provide the required cophasality in receivers of the discrete communication systems, there are special correction devices.

Given below is the classification of the discrete communication systems produced according to the most characteristic criteria: according to the form of the transmitted information, according to the method of the maintaining of cophasality of the distributors and according to the method of the use of a transmission cycle (reception).

According to the form of the transmitted information, the discrete communication systems are divided into data transmission systems and systems of the transmission of texts.

According to the method of the maintaining of the cophasality of distributors, all the discrete communication systems are divided into synchronous and start-stop.

The synchronous systems are called such systems whose distributors operate continuously independently of whether messages are transmitted or the system is found in quiescent conditions. Correction devices which provide the maintaining of the cophasal operation of the distributors continuously trace the phase relationship and if necessary eliminate the mismatch.

In start-stop systems the receiving and transmitting distributors are connected only with the transmission (reception) of the code combinations. In the absence of transmission the distributors do not operate.

The cophasality of the distributors is attained in that the receiving distributor is started immediately after the beginning of the operation of the transmitting distributor (start) and upon termination of one cycle is stopped (stop). Due to this the phase split between the distributors, which is stored toward the end of the cycle, is eliminated and at the beginning of the following cycle the cophasality of the distributors is provided by the stability of the drive.

According to the method of the use of a cycle of transmission (reception), the discrete communication systems are divided into single and repeated.

In single systems connected to the transmitting distributor is one storage element and connected with it is the coding

device (Fig. 7.1C). In single systems during one cycle of the transmitting distributor one code combination is transmitted.

If connected to the transmitting distributor are several storage elements, each of which will have their coding device, and during one cycle of the distributor to insure the transmission of several code combinations (according to one combination from each coding device), then we will have a multiple system. Figure 7.12 gives a block diagram of the transmitter of multiple communication system.

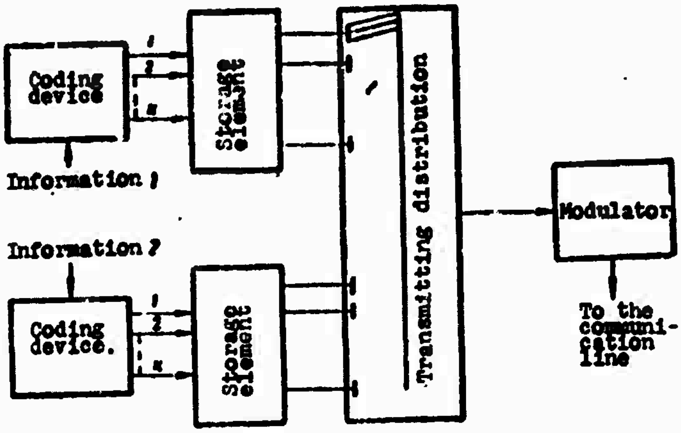


Fig. 7.12. Block diagram of a transmitter of the multiplex diode system of communication.

As can be seen from the diagram, the multiplicity of the system is determined by the quantity of sources of information (coding devices with the storage elements), between which the cycle of transmission (reception) is divided.

Multiple systems allow, on one hand, using more fully the transmission capacity of the line and on the other hand grant the independent communication channels to a larger number of correspondents. Multiple systems are occasionally referred to as the discrete communication systems with time-division multiplex.

Let us note that all the start-stop signals are single, while the synchronous systems can be single and repeated. It is necessary to add to this that all the discrete communication systems in the form of utilized elements are divided into electromechanical and electronic.

Electromechanical refers to systems built on contact elements (electromagnetic relays, disk or cam distributors, etc.).

Electronic systems are those which are constructed on the noncontact switching elements, irrespective of whether they are built around electron tubes, semiconductor devices, ferrites, and so on.

7.10. Layout and Operating Principles of One of the Existing Systems of Data Transmission to the Automatic Control System

In the foreign press there are reports about use in the USA of the equipment AN/TSQ-7 for the transmission of radar data along telephone channels in systems of the fire coordination of antiaircraft artillery, in particular, in the system "Missile Master."

The system AN/TSQ-7 provides data transmission on 48 aerial targets. By means of the system (in the rectangular coordinate system) the coordinates of targets making up the vectors of speeds of the targets in a horizontal plane (v_x, v_y, and also auxiliary information - the number of target, the characteristic of states of facilities and instruction are transmitted.

In the transmitter systems of the DC voltage, which reflect the transmitted information, are converted to numbers of the binary code being transmitted into line at a rate of 750 bits. Unity corresponds to the presence in the line of the signal of the carrier frequency of 1500 Hz, and zero - the absence of signal.

The length of each unit pulse (pause) is 1.33 μs. The transmission of the message about one target occupies 109.30 μs, while the transmission of data on 48 targets, consequently, occupies 5.24 s.

To account for the distance between the points of the reception and the transmission of information, the system provides for the reduction of coordinates to the single reference system.

A block diagram of the transmitter of the system is depicted in Fig. 7.13 and of the receiver - in Fig. 7.14.

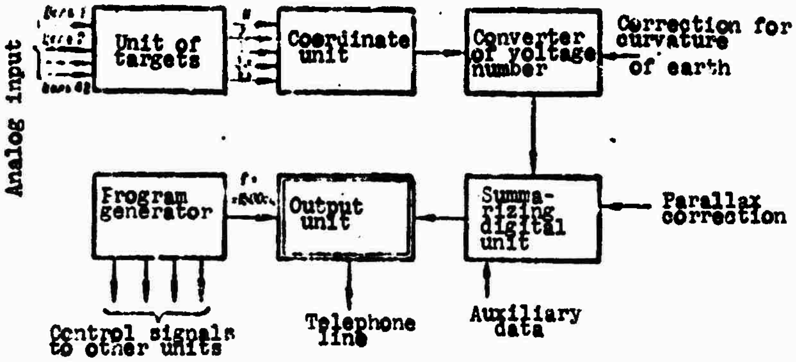


Fig. 7.13. Block diagram of the transmitter of system AN/TSQ-7.

Entering into the units of targets are voltages characterizing the coordinates and speeds of 48 targets, where they are temporarily stored. The unit retains the information about every target and transmits it to the coordinate unit in the sequence determined by the program generator. At each given moment consecutively

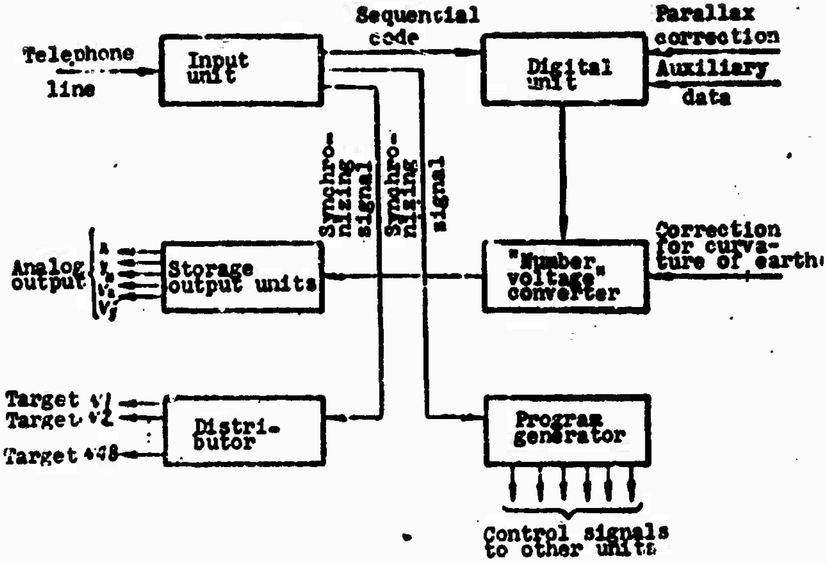


Fig. 7.14. Block diagram of the receiver of the system AN/TSQ-7.

the voltages characterizing the coordinates and speed of one target is transmitted. The coordinate unit, just as the unit of targets, has elements which serve for the memorization of the coordinates and speeds of one target for the period of serial transfer into the converter "voltage-number" of each of the voltages characterizing the coordinates and components of the vector of speed of one target.

In the converter the DC voltages are converted into a sequential digital code, and correcitons for the curvature of the earth are also introduced.

The converter operates in the following manner. The sawtooth-voltage oscillator continuously produces a linearly changing DC voltage. The rate of change in the voltage is maintained constant with high accuracy. By means of a comparison circuit this voltage is compared with the input voltage, which reflects one or another coordinate or velocity component. The moment of the equality of the voltages is fixed by the decimal counter, which is started simultaneously with the beginning of each period of a change in the sawtooth voltage. The counter generator generates pulses with a frequency of 100 Hz.

The quantity of pulses computed by the counter at the moment of the equality of the voltage of the converter and transmitted value will correspond to the decimal digital code of the transmitted value. In the digital unit the obtained decimal number is converted into a binary number.

The target position data and velocity components are transmitted also sequentially into the summing digital unit. Auxiliary data and the parallax correction between the point of standing and beginning of the single system of rectangular coordinates into the summing digital unit are introduced. The auxiliary data and parallax are introduced by means of a set of corresponding codes reflecting these data. The codes are typeset manually by means of telegraph keys, each of which has two positions corresponding to "zero" and "one."

From the summing unit finally the formed codes sequentially enter into the output unit. The latter oscillates at a frequency of 1500 Hz by means of a high-stability tuning-fork oscillator.

The codes of information, which are the sequence of ones and zeros, control the turning on and turning off of the signal being transmitted into the line.

The program generator produces control pulses with a repetition frequency of 750 Hz (1500:2) and feeds them to other units for synchronization of the fulfilled operations.

The total message consists of 82 time intervals with a length of 1.33 ms. The structure of the standard signal is shown in

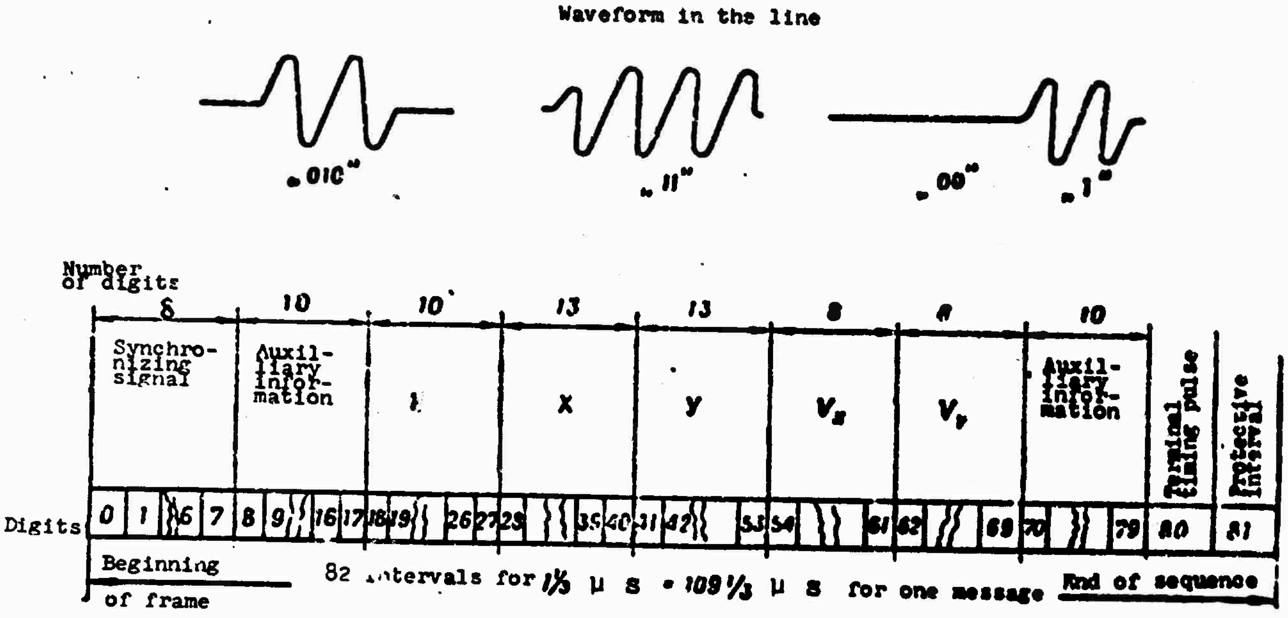


Fig. 7.15. Structure of the standard message.

Fig. 7.15. The first eight intervals are occupied by a synchronizing signal. The second ten intervals (bits after the synchronizing signal) are removed for the supplemental data and the 52 subsequent intervals (bits) are removed for the coordinate data, and of them: the first 10 - for height; the following two words of 13 digits each - for the transmission of coordinates X and Y; and two words of 8 digits each - for velocity components v_x and v_y.

Intervals of 70-79 are used for the transmission of the second word of auxiliary information.

The following then 80th interval (for the terminal timing pulse) is free provided the transmitted sequence is not the last for this target complex.

The last 81st interval (reading is conducted from the sequence) is protective and always free. After it the transmitter consecutively transmits analogous sequences - the information about the subsequent targets. At the end of the transmission of the message about the last target, into the 80th interval the "terminal synchronizing pulse" is introduced. Owing to this the receiver is synchronized with transmitter independently of the number of targets transmitted. The receiving system operates in a sequence opposite to the sequence of the operation of the transmitter. The signals entering into it are amplified and then are detected and divided. The selected pulses of information are transmitted sequentially into the digital unit. In the latter the appropriate parallax components are subtracted from values of coordinates X and Y. In the digital unit, auxiliary information is also separated. Pulses of the code of this information put the receiving relays in a position which reflects "zero" or "one." Each of the 20 digits of auxiliary information is represented by one relay. These relays hold the assigned position during the entire duration of the transmission of information on one target.

From the digital unit values of the target position data and velocity components enter into the "number-voltage" converter, which converts values from digital form into DC voltages. Here corrections for the curvature of the earth are considered.

The operating principle of the converter is explained by the diagram depicted on Fig. 7.16.

The DC voltages representing X, Y, H, and also v_x and v_y are consecutively connected to the storage and output unit, in which stored in separate circuits are the memories of the value of H, X, Y and v_x and v_y of each target during the time necessary for the serial transfer of these values into the storage unit. Upon the termination of the reception all five voltages simultaneously approach the output. Operations in the receiving system are synchronized by the program generator.

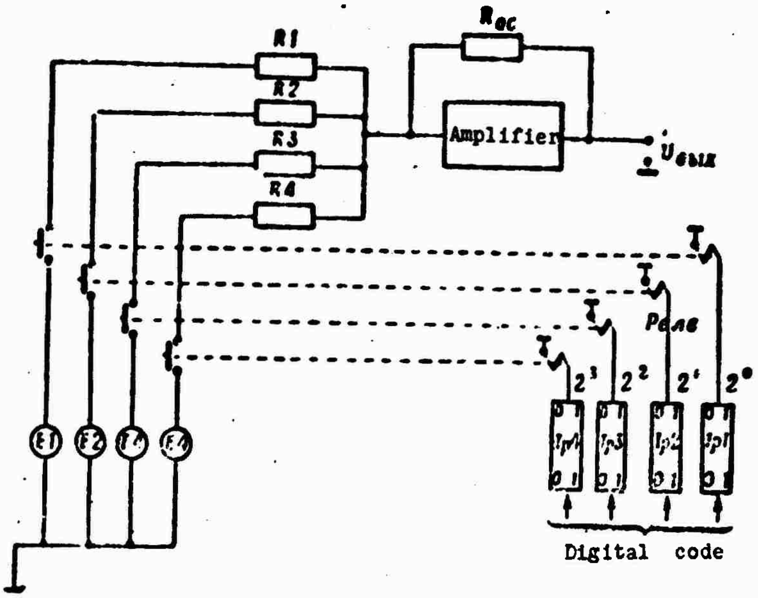


Fig. 7.16. "Number-voltage" converter.

The distributer is assembled on small-size relays similar to the analogous unit at the transmitter. This device makes it possible to relay the received data on each target along 48 separate lines serviced at the ends by the receiver.

The input signals, amplified in the input unit but still not detected, enter into the distributer, in which data on each of the targets are divided and directed along the appropriate lines toward the receivers.

CHAPTER 8

METHODS AND MEANS OF THE DISPLAY OF INFORMATION

8.1. Methods of Visual Display of Information

Features of the conducting of contemporary combat operations (increased dynamicity, increase in the speeds, ranges and destructive power of the weapon) lead to the fact that at the command posts there is processed such an information flow which one person often cannot comprehend in those short time intervals during which it is available in the process of the control of combat. In connection with this the selection of rational forms of the bringing of information processed by the computer to the appropriate persons is very important.

Since of all the sense organs of man the maximum transmitting capacity is processed by the visual channel, the primary value is acquired by methods of the visual representation of information. The display units of information in automated control systems basically reproduce the visual information.

The information reproducible at the command posts of automated control systems can differ by the position and orientation of the characters, the gradation of brightness or tones of the monochromatic image, by color and its intensity, by form and dimensions, and also by the derivatives of these parameters, for example, by the flickering movement or storage of the image. Of the many methods of visual representation of information in the automated control systems, at the present time the most widespread use is the alphanumeric method, the method of the representation of information by characters of an arbitrary form, the method of the delineation of the lines, and the method of the designation of areas.

With the alphanumeric method of the representation of information, the data are reproduced by means of letters and digits represented in the form of tables or text. This method is the most common and is quite effective. Investigations show that the operator's reaction when using an alphabet of digits and letters is more active than when using more complex alphabets.

The use of characters of arbitrary forms is also a very effective method of the expression of complex thoughts, subjects or events. A shortcoming of the method should be considered the the comparatively small number of different characters suitable for visual representation due to the limitedness of storage of the operator and difficulty of the designing of characters with high mnemonic value.

Some forms of information are difficult to represent on visual displays by means of the alphanumeric signs or characters of an arbitrary form. Thus, air routes, highways, topographic contours, the routes of movement, radii of zones of action, etc., are best of all represented by the method of the delineation of lines. Furthermore, this method is rather simply implemented when data enter from the output of the computer.

Representation of areas is necessary for designation of the basins, combat areas and sectors contaminated by toxic or radioactive materials of the sectors. The areas can be reproduced by the method of the delineation of the lines, by the use of characters and shading, and also by the compilation of remarks. The coloring or blackout of areas are promising; however, at the present time these are difficult to attain.

The evaluation of methods of the visual representation of information depends on such factors as psychological features of the activity of the operator receiving the information (for example on the possibility of the operator of distinguishing the gradations of brightness or color), on the relative compatibility of methods of coding and on a whole series of other reasons.

At the command posts of the automated control systems, usually all the enumerated methods of the representation of information in different combinations are applied.

Graphical information from the computer is displayed by means of different devices which, according to the method of their technical implementation can be divided into the following groups:

- electromechanical devices of the displaying of graphical information [UVGI] (УВГИ);
- electronic UVGI;
- electroluminescent UVGI;
- UVGI based on microfilms;
- photochemical and electrochemical UVGI;
- combined UVGI.

In electromechanical UVGI digital signals from the computer or another device which transmits digital information, after the conversion into analog form and amplification, are fed to the servodrive of the mechanical drawing device.

In electronic UVGI graphical information is displayed on the screen by electronbeam tubes [ELT] (ЭЛТ).

The basic part of the electroluminescent UVGI is the matrix (screen) consisting of separate cells with an electroluminescent cover. The luminescence of the separate matrix cells forms on screen the necessary alphanumeric and graphical information.

In the UVGI built on the basis of microfilms, the major carrier of information is the microfilm (or diapositive) on which the information is directly fixed in graphical form.

In photochemical and electrochemical UVGI the image is obtained because of the reversible or irreversible chemical reactions occurring in the sensitive layer of the electrochemical paper, photographic plate, and so on. In the combined UVGI the different principles described above are combined.

8.2. Methods of the Formation of Elements of The Image in the Electronic Devices of Displaying Graphical Information

For the visual representation of information presented in digital code, the digits must be translated into characters, lines or contours and in a definite order placed them on the screen of the device of visual representation.

In general form the block diagram of the device of visual representation of information takes the form indicated on Fig. 8.1.

In the electronic devices of the display of graphical information, for the formation of the characters the most widely used are the following methods:

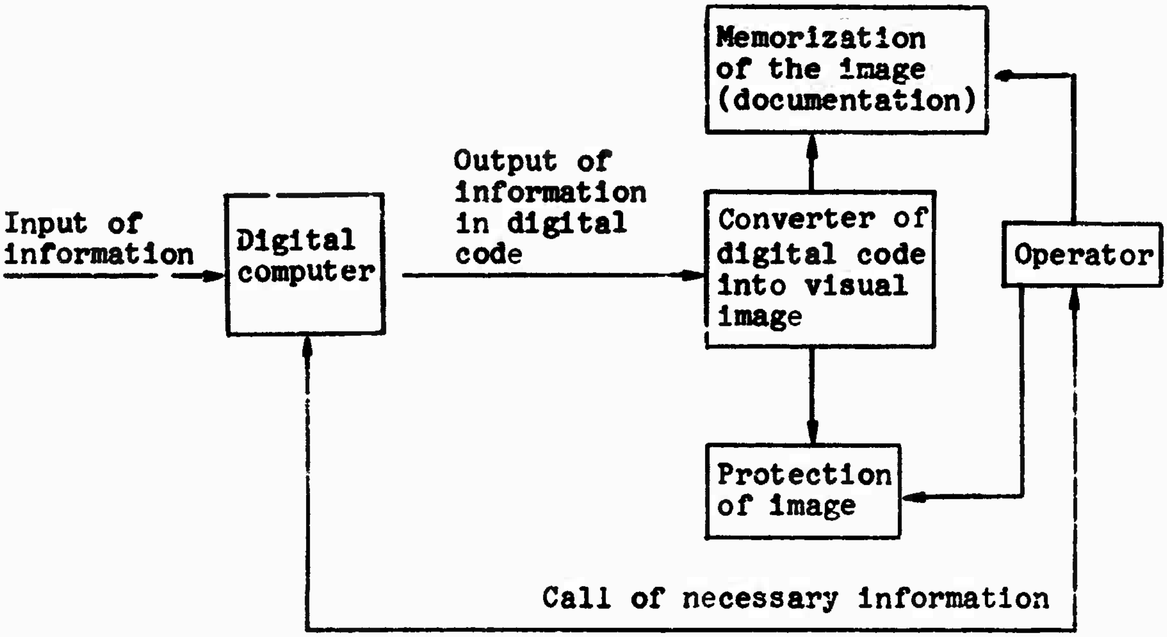

Memorization of the image (documentation)
Output of information in digital code
Input of information
Digital computer
Converter of digital code into visual image
Operator
Protection of image
Call of necessary information

Fig. 8.1. Block diagram of the device of visual representation of information.

- the method of the dot image, which consists in the formation of one or another character from isolated dots;

- the method of scanning of the CRT (Cathode Ray Tube) beam; in this case for each character its form of scan by means of a grid of magnetic cores is created;

- the method of Lissajous figures, which uses scanning of the image and its illumination along the sign trace;

- the method of the figure beam, which consists in the fact that assigned to the transverse areas of the beam of the cathode-ray tube is the form of that character in the form of which it is necessary to represent information.

All these methods are united by the general principles of the construction of the block diagrams. Each of the shcemes must have a unit of identification of the character, a unit for positioning the character on the screen, a memory unit and a cathode-ray tube.

Figure 8.2 depicts a block diagram of the device for the information of characters by the method of the dot image. When using the method of the dot image the characters are formed from separate dots of the contour corresponding to one or another character. Each dot in turn is formed by deflection of the beam by the cathode-ray tube along axes X and Y by the assigned magnitude (at the necessary place on the screen). Then the cutoff voltage is removed from the tube, and the dot is projected on the screen.

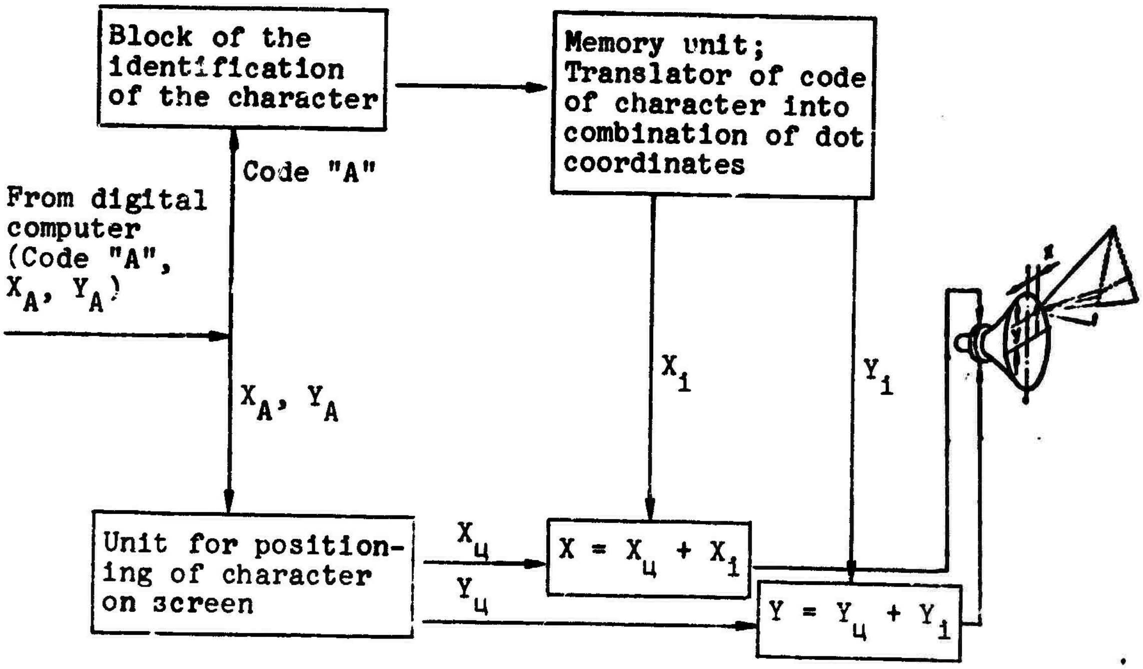

Block of the identification of the character
Memory unit; Translator of code of character into combination of dot coordinates
Code "A"
From digital computer (Code "A", X_A, Y_A)
X_A, Y_A
X_1
Y_1
Unit for positioning of character on screen
X_4
Y_4
$X = X_4 + X_1$
$Y = Y_4 + Y_1$
X
Y

Fig. 8.2. Block diagram of the device for the information of characters by the method of the dot image.

Each character is formed from the configuration of dots which are defined and constant for this character. The number of dots, the coordinates of divergences of each dot along the axes, and the relative position of the dots for each character are stored in the memory unit. The memory units can be constructed in the form of circuits containing resistors or in the form of ferrite matrices or a magnetic drum.

As a result of the data processing, the digital computer issues a combination of codes of characters. Each code of the character is characterized by a conditional subscript and place of position of the screen (for example, the code of the character "A": and the code of the position of this character on the screen - X_A and Y_A). The code of the character enters into the unit of the identification of the character, where according to its subscript the coordinates of the position of dots in the memory unit making up this character are determined. Having obtained the signal about the output of the image of the character, the memory unit sequentially issues coordinates of each point relative to the center of contour (X_1, Y_1). According to the code of the position of the character on the screen (X_A, Y_A) the unit of the position of the character produces coordinates of the position of the center of the character on the screen (X_4, Y_4), which are then added to the coordinates of dots (X_1, Y_1), and the resulting divergences (X and Y) are fed to the cathode-ray tube screen. As a result of the sequential transfer of dots at the assigned place of the screen, the character issued by the computer is represented.

The method of the formation of the character by means of scanning is used with rasters (forms of the scanning of characters), which are formed with the help of a grid of magnetic cores or an optical stencil (masks). The formation of the characters with the help of a grid of magnetic cores is clarified in Fig. 8.3a. The ferrite cores of the grid are connected in different

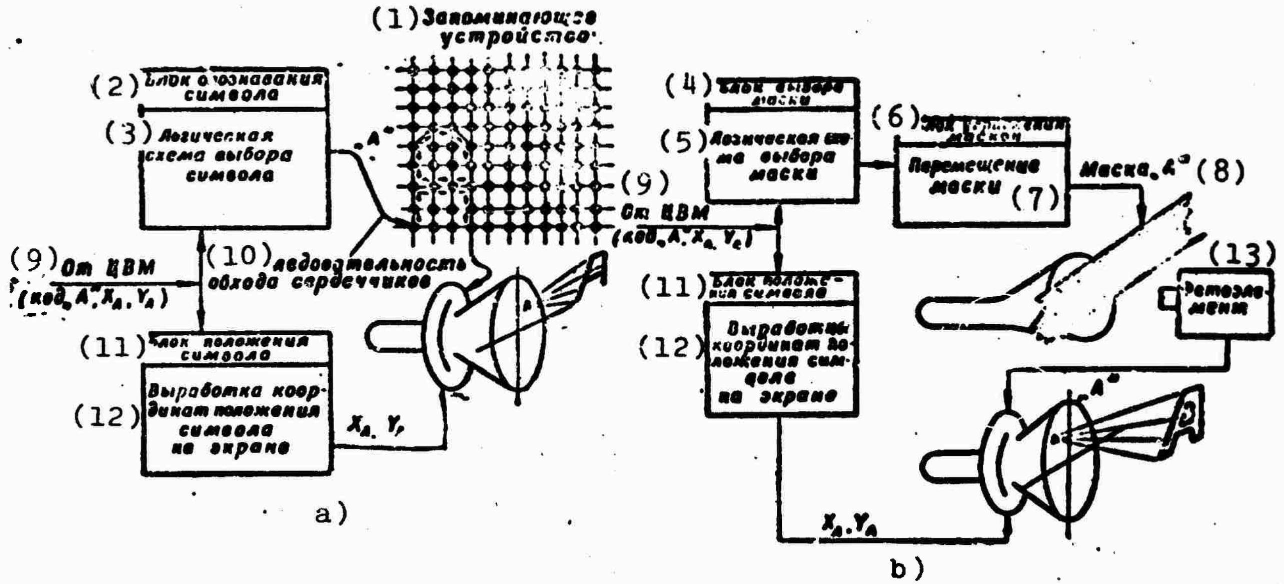

(1) Запоминающее устройство
(2) Блок опознавания символа
(3) Логическая схема выбора символа
(4) Блок выбора маски
(5) Логическая схема выбора маски
(6) Блок управления маской
(7) Перемещение маски
(8) Маска „A"
(9) От ЦВМ (код „A", X_A, Y_A)
(10) Последовательность обхода сердечников
(11) Блок положения символа
(12) Выработка координат положения символа на экране
(13) Фотоэлемент
X_A, Y_A
a)
b)

Fig. 8.3. Diagram of the formation of the characters: a) by the scanning method; b) by the optical mask method.
KEY: (1) Memory device; (2) Unit of the identification of the character; (3) Logic circuit of the selection of the character; (4) Unit of the selection of the mask; (5) Logic circuit of the selection of the mask: (6) Mask controls unit; (7) Movement of mask; (8) Mask "A"; (9) From computer (code "A," X_A, Y_A);
(10) Sequence of the circuit of cores; (11) Unit of position of the character;
(12) Development of coordinates of position of character on the screen;
(13) Photocell.

combinations, forming geometric outlines of one or another character. The computer issues, as in the preceding case, the code of character (A) and the code of the position of the character on the screen. The code of the character enters into the unit of the identification of the character, where the order of the circuit of the cores in the grid corresponding to each character is stored. The ferrite grid controls the beam of the scan of the cathode-ray tube. As a result the beam of the scan will also draw the letter A on the tube screen. Simultaneously in the unit of the position of the character voltages are produced which are proportional to the coordinates of dot of the position of the center of the character on the screen, in consequence of which the letter A will be depicted in the corresponding place on the screen.

Images of the characters are formed somewhat differently by the method of scanning with the help of an optical stencil (mask, Fig. 8.3b). The display unit contains an additional cathode-ray tube for the development of the raster of the signal. Situated between this tube and the photocell are masks of all the characters being used in this display unit. The masks are prepared from a material opaque for light. They have figure slits corresponding to the geometric outlines of the characters. The code of character enters into the unit of the selection of the mask. As a result of the function of the unit, the mask which corresponds to the character formed is located between the tube and the photocell. The beam of the raster will luminesce the character through the mask. The photocell controls the scanning of the beam of the main tube, on the screen of which this character is represented. The place of the representation of the character on the tube screen corresponds to the voltages produced in the unit of the position of the character.

For the formation of characters by the method of Lissajous figures, an oscillator producing sinusoidal oscillations of

different frequencies is used. From these frequencies outlines of different characters are combined by means of the limitation and selection of moments of their inclusion. The simplest principle of the writing of digits by means of Lissajous figures is explained by Fig. 8.4.

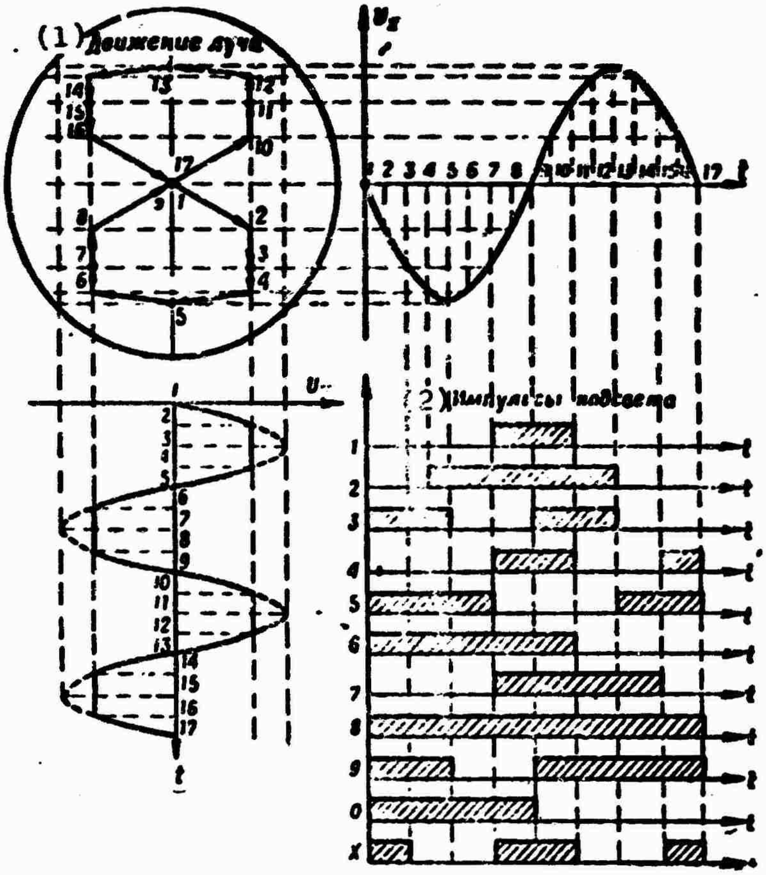


Fig. 8.4. Explanation of the principle of the writing of digits by means of Lissajous figures. KEY: (1) Motion of ray; (2) Illumination pulses.

Used for the writing of the digits is one of Lissajous figures, which is obtained with beam deflection by two sine voltages with frequencies of which one is twice higher than the other. The figure being formed here reminds one of a figure 8. The digits are obtained by means of the illumination of the beam at the appropriate moments of time when it follows along the profile of the eight.

With such a method of the writing of digits, the control consists in the selection of conductions of illumination and of the pulse shaping of the illumination.

There are other varieties of the image method of digits by means of Lissajous figures. As the example which explains one of such varieties, let us examine the shaping of number 4 (Fig. 8.5).

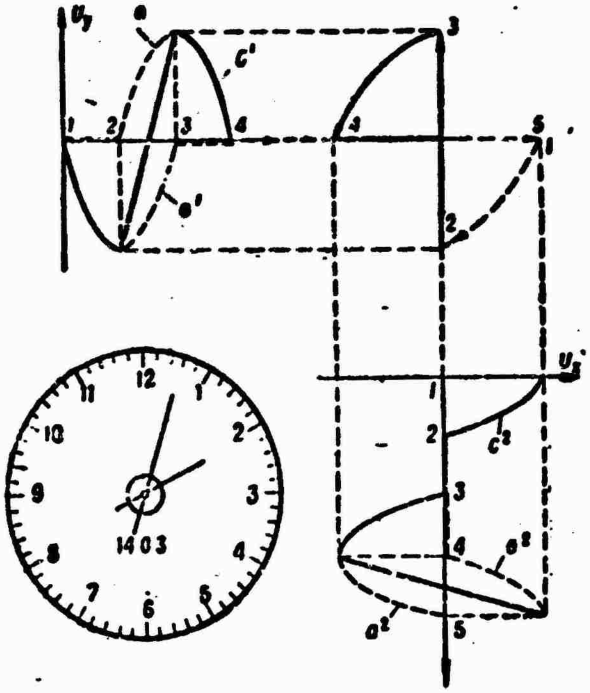


Fig. 8.5. Display of the dial of clocks obtained on the tube screen and an example of the shaping of voltages for representation of number 4.

The voltage of vertical deflection U_y in sector 2-3 is obtained by means of the addition of the positive half-wave of the sine voltage (a^1) with the negative half-wave (b^1) shifted

90° in phase with respect to the first. Solid line shows the resulting form.

The voltage U_x of horizontal deflection is obtained from the same sine voltage leading in phase the voltage of vertical deflection by 90° (c^2). In this case the negative half-wave of the voltage U_x is shifted with a lag relative to the positive half-wave for a certain time. On sector 4-5 the voltage a^2, being added to the positive half-wave, forms a figure similar to the voltage of vertical deflection. As a result of the action of voltages U_x and U_y on the cathode-ray tube screen, the sign shown in the upper right part of Fig. 8.5 is drawn. In order to obtain the Fig. 4, the dashed part of the sign is blanked. Using different combinations of the control voltages U_x and U_y, it is possible to obtain any display (for example, a model of the dial of clocks).

A characteristic feature of the enumerated methods of the formation of the characters is the fact that in them the electron beam of the tube is the writing element being used as a pencil.

In special sign cathode-ray tubes characters are written on the screens by means of an electron beam, to the cross section of which is assigned the shape of that sign which must be recorded. The electron beam seemingly print on the screen the necessary signs, unlike the devices with ordinary tubes where the recording of the signs requires the appropriate complex circuits of scanning and control.

In relationship to the clearness of the signs, the speed of the display and convenience of their use in conjunction with the computer, the sign cathode-ray tubes exceed all the remaining types of tubes intended for that purpose. As an example let us examine a sign cathode-ray tube of the Charactron type (Fig 8.6a).

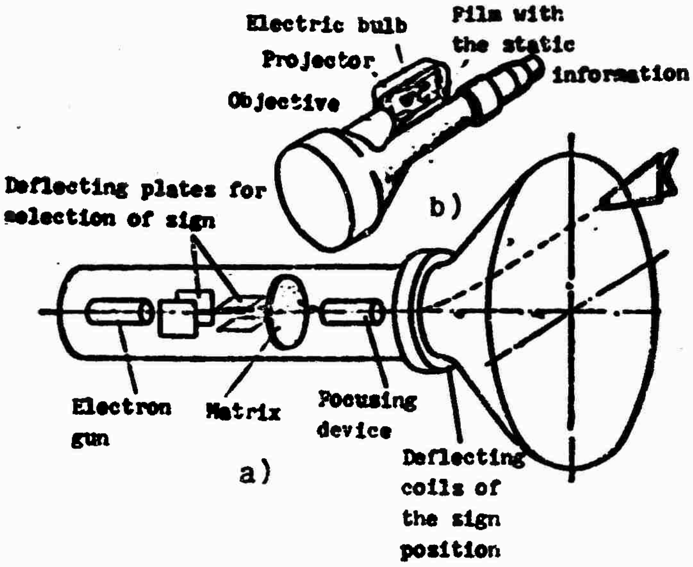

Electric bulb
Film with the static information
Projector
Objective
Deflecting plates for selection of sign
b)
Electron gun
Matrix
Focusing device
a)
Deflecting coils of the sign position

Fig. 8.6. Diagram of a device of a sign cathode-ray tube of the Charactron type: a) single-beam; b) with window in rear wall.

In such tubes installed between the screen and the electron gun is a matrix with a large number of small shaped holes. By directing the beam through these openings, it is possible to project on the screen the necessary combination of numbers, letters and symbols. For this at first the electron beam is deflected in such a way that it passes through the defined holes of the matrix and forms the necessary image. Then the beam is deflected for a second time in order that the formed conventional image would enter into the necessary place on the screen.

The signs of the matrix and their position on screen for display are selected by signals from the computer fed to the deflecting plates of selection of the sign and to the deflecting coil of the position of the sign. By alternating in a definite order the deflecting voltages on the screen of the charactron, it is possible to apply the combat situation and another information. Because of the high rate of movement of the beam and the presence of a sufficient screen afterglow, all signs on the screen are visible simultaneously.

The specific element of the Charactron is the matrix, and it determines the type and total amount of the signs used. Usually the matrix (Fig. 8.7) is made of metal with a thickness of hundredths of a millimeter, and area of the matrix is about 1.5 mm^2; arranged on it is not less than 60 signs, and the height of each sign is about 0.3 mm.

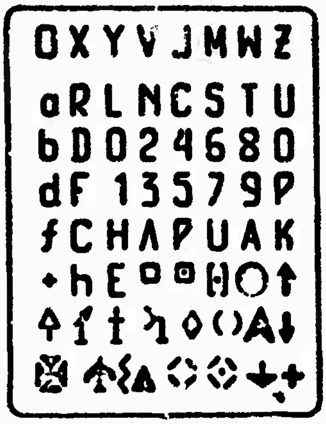


Fig. 8.7. Matrix of the Charactron S19K.

In the sign cathode-ray tubes intended for operation in systems with the use of circular-scan radar on the matrix, besides the signs there is a circular opening which makes it possible to pass the entire electron beam and then focus it on the screen just as in ordinary cathode-ray tubes. This makes it possible to use a sign tube also as a plan position indicator.

Besides the ordinary (single-beam) tubes of the Charactron type, double-beam tubes and also tubes with a window in the rear wall are manufactured. Through this window, with the help of a projector, static information is luminesced (Fig. 8.6b).

On screens of Charactrons used in air defense systems, the so-called "logbooks" of targets are used as characteristics of aerial targets.

The logbook of the target (Fig. 8.8) can be expressed by a group of several groups and digits. The center of each group determines the target position, and the pointer near the logbook determines the direction of its motion. The logbook is moved on the tube screen in accordance with the motion of the target. Figure 8.9 shows the screen of a Charactron on which pipes of targets with logbooks are visible.

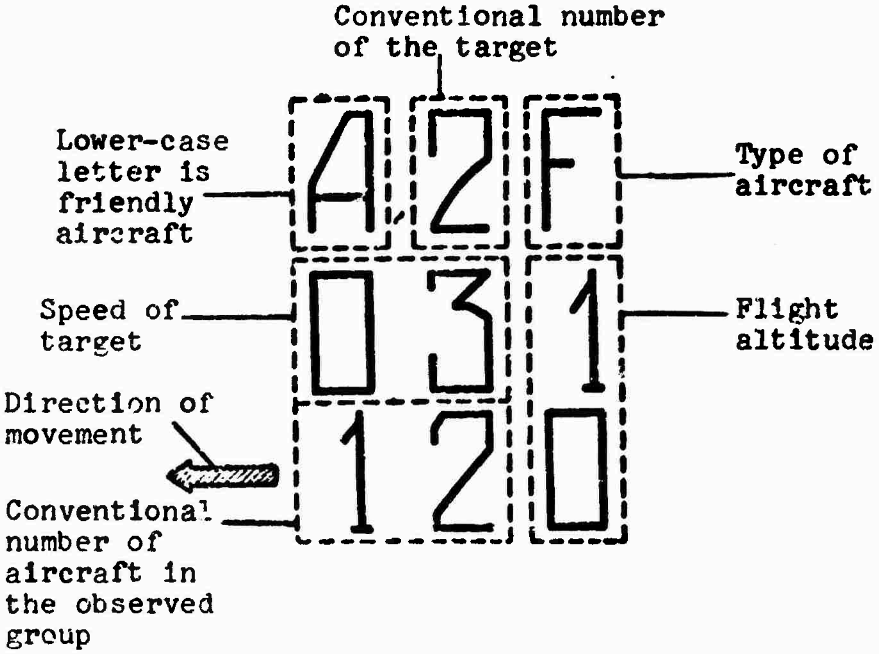

Conventional number of the target
Lower-case letter is friendly aircraft
A
2
F
Type of aircraft
Speed of target
0
3
1
Flight altitude
Direction of movement
1
2
0
Conventional number of aircraft in the observed group

Fig. 8.8. Logbook of a target.

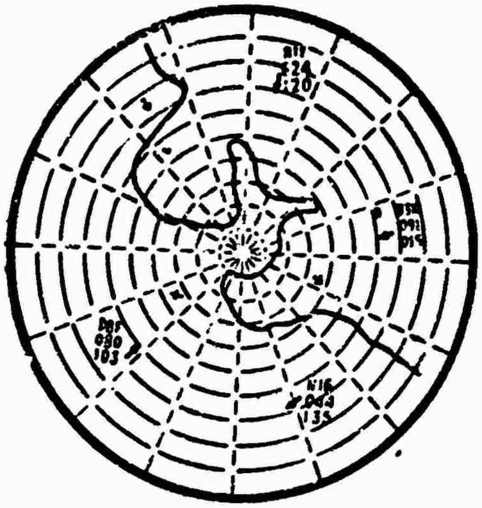


Fig. 8.9. Display of aerial situation on the screen of the Charactron.

Besides Charactrons, for the reproduction of an aerial situation a storage tube of the Typotron type can be used, and it represents a variety of the Charactron with a storage of information. In the Typotron the beam also passes through the matrix and two systems of deflecting plates. However, the information is recorded not on the luminescent screen but on a dielectric storage target, on which it can be stored for a very long time. The target is a small-structure grid which is located in parallel to the screen and several centimeters from it and is covered on the inside by a dielectric layer. If necessary the entire information from the target can be erased by a change in the supply voltage on one of the electrodes of the storage device. The Typotron and basic elements of its electron-optic system are given on Fig. 8.10.

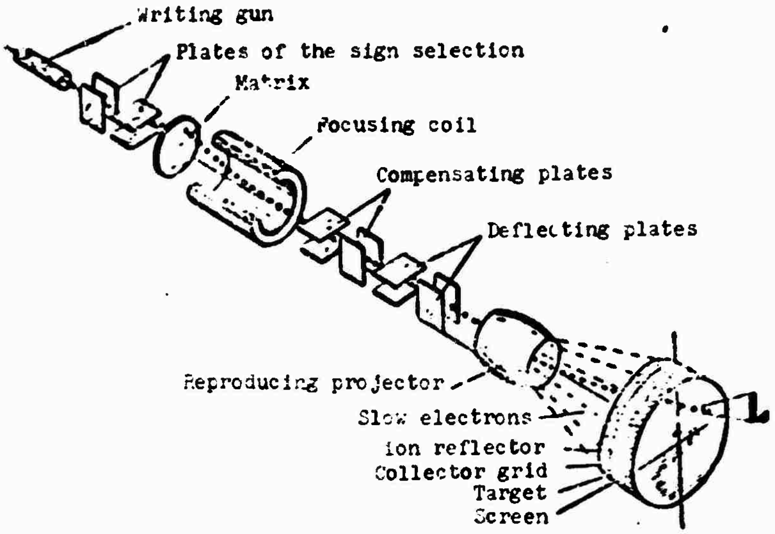


Fig. 8.10. Typotron and basic elements of its electron-optic system.

The operating principle of the Typotron can be explained in the following manner. Electrons of writing beam, which carry the image of the sign, bombard the surface of the target, which is the basic storage device element of the tube. The electrons knock out secondary electrons from the target surface. With the imparting to the beam of speed, the coefficient of secondary emission of the target becomes greater than one, and therefore the target at places undergoing bombarding by electrons is charged positively approximately up to the potential of the collector grid, which is in front of the target. Thus, a potential relief in the shape of the appropriate sign is formed on the target.

Besides the writing gun, which operates for a short time only at the moment of the writing, in the tube there is a reproducing gun with an unfocused beam. This gun creates the wide uniform beam of the low-velocity electrons irradiating the entire target.

The cathode of the reproducing gun has a potential approximately equal to the potential of the metallic backing of the target and is usually connected with the housing.

The flow of slow electrons, which is directed through the collector grid on sectors where there was nothing written, encounters the decelerating field of the target and is reflected from it. Therefore, at places of the target where there is no recording electrons do not pass through the dielectric. At places where there is a recording of the sign, electrons pass without difficulty through the grid cells and under the action of the accelerating field strike the screen, forming an image of a sign.

Information recorded on a Typotron in conditions of storage can be retained for an arbitrarily long time. The image on the screen can be erased, having lowered the potential of the collector grid. In this case the value of the potential of the surface of the dielectric will be lower than the critical, and under the action of the flow of slow electrons it is reduced to zero. A new recording is possible only after the establishing of nominal values of voltages on all electrodes of the tube.

Installed in front of the collector grid is a grid of the ion reflector, to which a positive potential is fed. The purpose of the grid is to block the target against the entering of positive ions, which are formed in envelope as a result of the ionization of residual gases. On the walls of the envelope between the reproducing gun of slow electrons and the screen there is applied a layer of Aquadag (3rd anode), to which a positive potential is fed. Different potentials of the 2nd and 3rd anodes create the lens - collimator, the purpose of which is the conversion of the diverging flow of slow electrons of the reproducing gun into the flow perpendicular to the surface of the target. Such a conversion is necessary for the prevention of distortion of signs during their transmission from the target to the screen.

The average time necessary for the recording of one sign is approximately 40 μs. For the erasure of the written image, approximately 50 μs are necessary. The condiderable erasing time of the image impedes the use of the Typotron.

Signs reproducible on the screen of the Typotron are so bright that there is a possibility of using the Typotron for observation of information written on it even in the presence of external illumination.

The basic use of the Typotron is as an electronic signal panel instead of different light tables. Furthermore, Typotrons with dimensions of the screens of up to 500 mm can be used as indicators in systems where the display of a large volume of information with its one-shot output is necessary.

Usually the matrices of electronic UVGI have 64 or 128 different characters and a speed of generation of 25,000 to 200,000 characters per second. For screens of the tubes phosphorous with a small afterglow is usually used. To avoid noticeable flickering the rate of the exchange of the frames should not be less than 40 per second, and therefore practically with the indicated speeds of generation on the screen it is possible to illuminate not more than 600-5000 characters.

Many electronic UVGI are equipped with a buffer storage the basic purpose of which is the freeing of the computer from operation in the exchange of images. For example, if an image must be changed 40 times per second in the presence of 2000 characters on the screen, then each second entering into the devices which do not have buffer storage should be about 80,000 information words. Equipment of the buffer storage is located between the output of the computer and the display unit. The equipment stores the information coming from the computer, and this information can be used repeatedly. The use of the buffer storage makes it possible to reduce the information flow of the computer by 1000 and more times.

In all the examined sign cathode-ray tubes there are functional elements which should be affected by control signals for the obtaining of the image of the signs. Such elements are:

- device for shaping the electron beam and the control of it;

- system of the selection of signs on the matrix;

- address system for the arrangement of signs on the screen.

Since the basic purpose of such tubes is their operation in conjunction with the computer or other devices which produce information in binary code, then the control circuit should contain code registers in which is the entered information is stored and converters for converting digital values of the code to continuous voltages or currents proportional to it.

In this case the control circuit should have a device which would determine the sequence of the action of control voltages on functional elements of the tube.

Figure 8.11 depicts one of the possible variants of the control circuit of the Charactron.

Entering into the control circuit from the source of information which produces it in the form of a parallel binary code are the codes which locate the beam on the tube screen and determine the selection of the required signs of the matrix and also the control signals which are necessary for the communtation of the separate units of the control citcuit. One control signal determines the rate and sequence of operation of the control circuit and the other is intended for the return of all the registers to the initial position before the beginning of the next cycle of the display of information. The information enters into the control circuit in a definite sequence. At first codes determining the position of the beam on the screen enter into the commutator. Then codes of the position of the beam are converted into continuous values of currents or the voltages and through push-pull amplifiers act on the address system of the tube, establishing the beam at an assigned place on the screen. In this case the tube is cutoff.

After the passage of the codes of the position of the beam codes of signs of the matrix enter into the commutator. In this

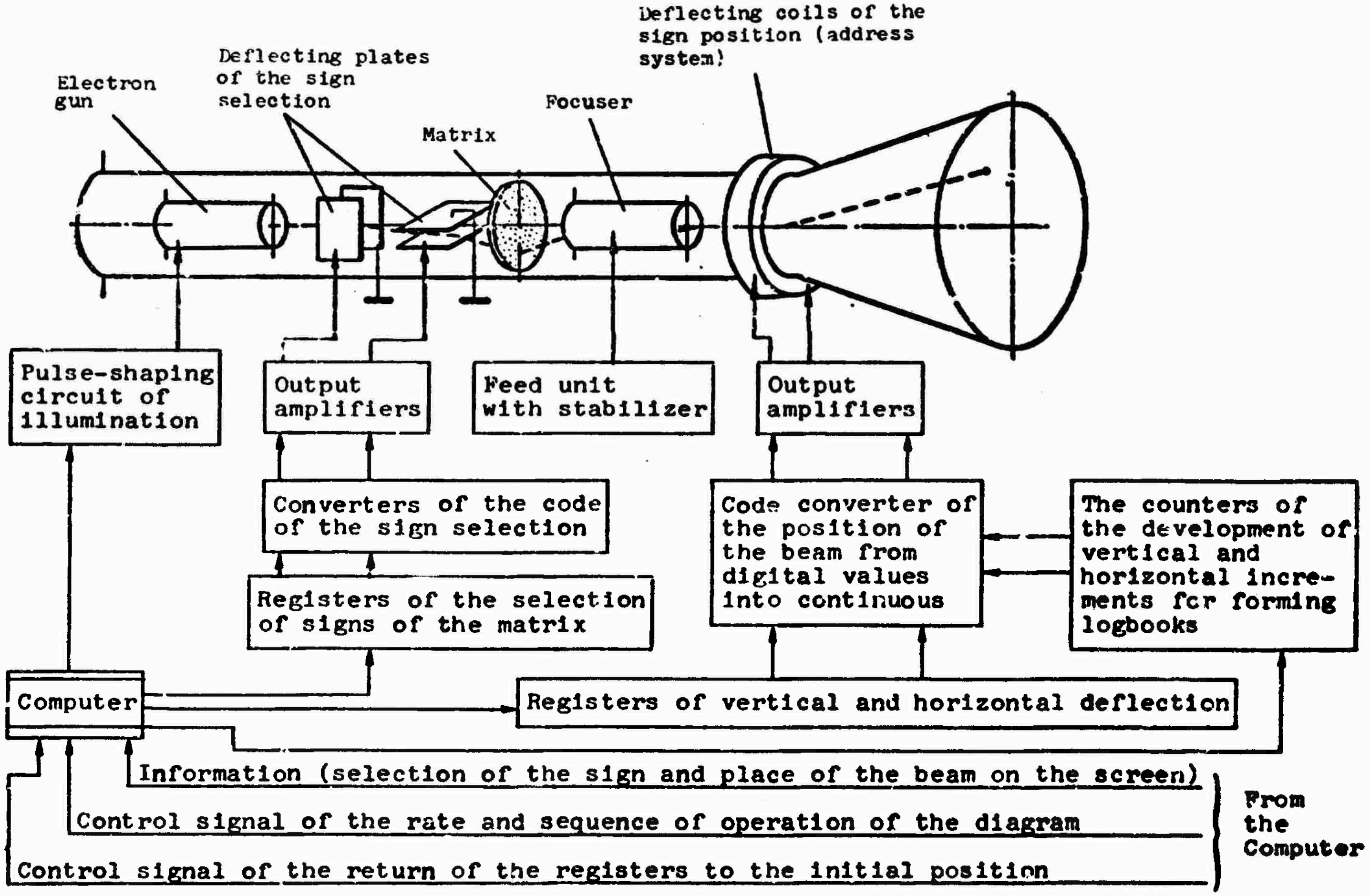

Deflecting coils of the sign position (address system)
Electron gun
Deflecting plates of the sign selection
Focuser
Matrix
Pulse-shaping circuit of illumination
Output amplifiers
Feed unit with stabilizer
Output amplifiers
Converters of the code of the sign selection
Code converter of the position of the beam from digital values into continuous
The counters of the development of vertical and horizontal increments for forming logbooks
Registers of the selection of signs of the matrix
Computer
Registers of vertical and horizontal deflection
Information (selection of the sign and place of the beam on the screen)
Control signal of the rate and sequence of operation of the diagram
Control signal of the return of the registers to the initial position
From the Computer

Fig. 8.11. Control diagram of a Charactron.

case codes of the position of one sign of the matrix on the vertical and horizontal enter simultaneously. The entered codes from outputs of the commutator act on the registers of the selection of signs of the matrix and on the converters.

The continuous voltages, which are proportional to the values of the codes, proceed through push-pull amplifiers from the converters to the deflecting plates of the tubes, determining the selection of one or another sign on the matrix.

In order that on the screen of the Charactron there would not be observed a trace of the changeover of the beam from one sign of the matrix to the other, each code of the sign is finished by an extinguishing pulse.

For the clear operation of sign cathode-ray tube; the stability of the current in the focusing coil is of special importance. For this purpose high-stability stabilizers are used.

8.3. Electromechanical Devices of the Output of Graphical Information

Electromechanical UVGI refer to automatic drawing machines and also special-purpose devices for the derivation of information on a large screen. Usually these devices are used for the demonstration of variable data against an invariable background (for example, the transfer of aerial targets on a geographical map). In this case the background is projected on a screen by means of an individual projector. If necessary the background can be changed. An example of such a device is the dynamic electromechanical projector of the firm Temco Electronics (USA), a diagrammatic representation of which is given on Fig. 8.12.

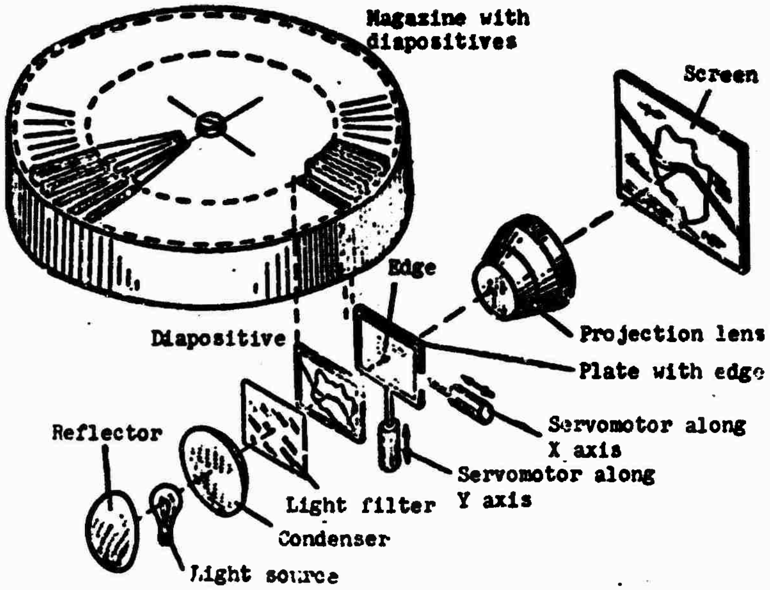


Fig. 8.12. Diagram of a dynamic electromechanical projector.

Fed from a drum-type magazine to the operating position is a diapositive. It is arranged in paralled to a glass plate, in center of which an edge is fastened. First the plate is moved so that the edge touches the emulsion cover of the diapositive. The plate with the edge, moreover, by means of a servomechanism can be moved along the X and Y axes. When control voltages are applied to the servomechanism, the plate with the edge is moved and the edge draws a trace on the cover. As a result on screen the background and traced line are projected. The diapositive use can be returned to the drum, and in its place a new one is installed.

For the displaying on the screens of information about the aerial situation in the scale of the entire country, at command posts of air defense of the USA the electron-optic display system Iconorama is used.

This system consists of the following elements:

- large-wall screen;
- several projectors with recording devices;
- autonomous projector plane tables;
- control panel of the projectors;
- control boards and reproducing devices for the multiplication and documentation of the recorded information.

The system is multichannel. For the display on a large screen of a complex situation simultaneously several projector devices are used. The basic element of the system is a projector device within which is built in the miniature recording mechanism (Fig. 8.13) consisting of an automatically rewinded transparent film covered with an opaque metallic layer and a very thin scratching needle rigidly fastened to a transparent holder. The holder with the needle is moved along the X and Y axes by two servo systems in accordance with the entering signals. Servomotors of the servo systems are connected with the holder of the needle by precise screw transmissions.

On the termination of the recording or upon the exchange of the sequence, the needle is removed to the side by means of a solenoid. The recording mechanism is placed within the optical system with a colored light filter.

The dimension of the projection reproducible by one projector is 2.5 × 2.5 m. Supplemental data in the form of cards, coordinate grids, alphanumeric designations, etc., can be superimposed on the obtained image by means of other projectors.

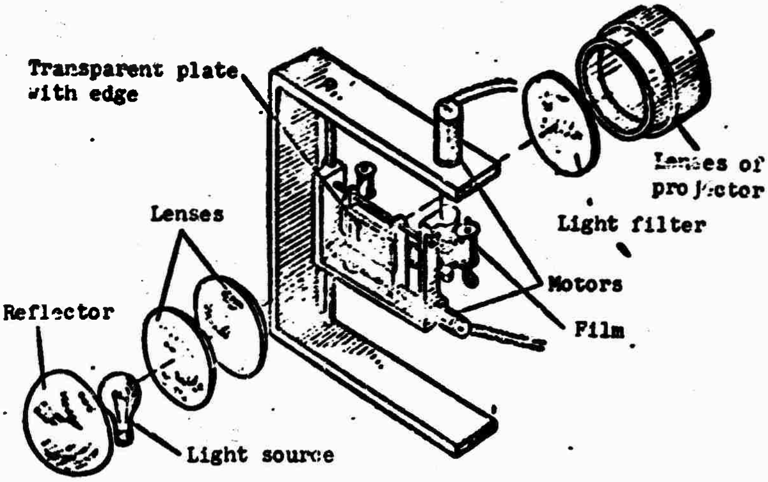


Fig. 8.13. Diagram of a projector device of the Iconorama system (variant with the film being moved).

There is the possibility of displaying the aerial situation in three projections, which is one of the advantages of Iconorama. A volumetric image can be obtained because of the use of projectors with a steroscopic optical system.

8.4. Electroluminescent Devices for the Output of Graphical Information

Electroluminescence is connected with the emission of light by luminescent materials (phosphorus). In electroluminescent devices for the output of graphical information the phosphorus is activated (glows) in the electric field created by the alternating voltage. The phosphorus is included between two plates of a condenser and is dielectric. One of the plates - a transparent electro-conductive material - is applied to the glass or film and passes light being radiated by the phosphorus. The second plate of the condenser is a layer of aluminum applied to the phosphorus; it serves not only for the application of the alternating electric field and light reflection, but can be approximately divided into separate elements isolated from each other, from which different characters are compiled.

One of similar devices created by the firm RCA is intended for the display of straight lines, letters and numbers on electroluminescent panels. The electrical circuits of the panels are comparatively simple in connection with the fact that in each element of the panel there is "built-in" storage. Thus, one coding device can service several panels consecutively. Each of them glows until it is discharged by means of the cutting off the supply of driving voltage. After this the brightness practically falls instantly; however, for the repeated turning on of any element of the panel about a fourth of a second should pass. The panel consists of a large number of moduli, each of which in turn consists of a number of segments. From the segments it is possible to combine different characters (Fig. 8.14 shows the modulus and the characters obtained by it).

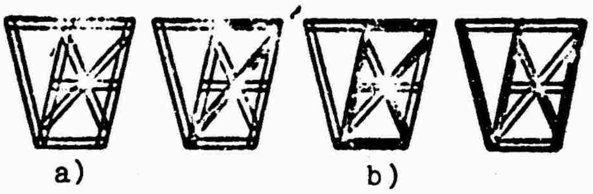


Fig. 8.14. Diagram of the modulus of electroluminescent panel a) and examples of characters obtained on its base b).

An electric diagram of one segment of the panel is depicted on Fig. 8.15.

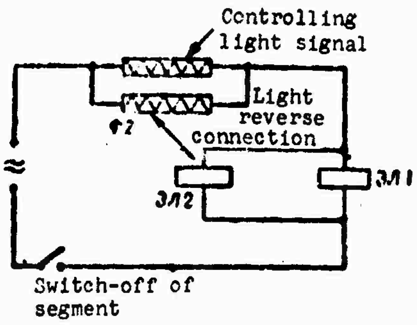
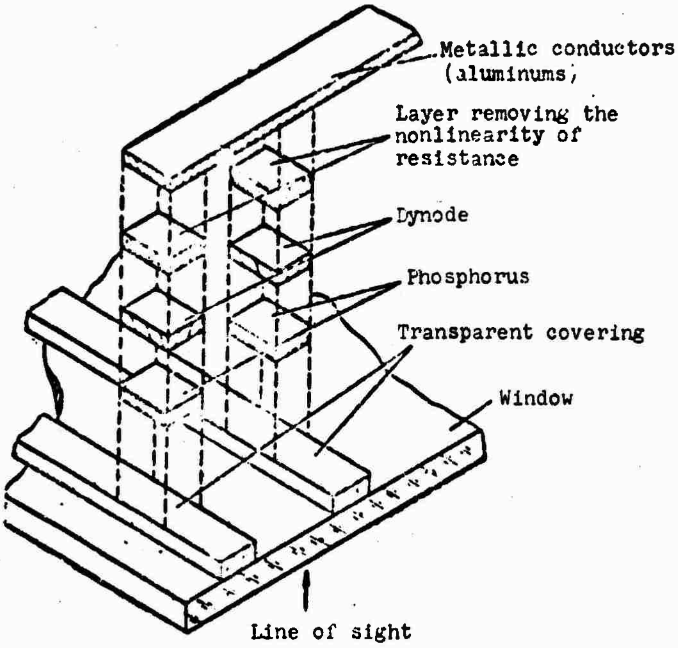


Fig. 8.15. Electrical circuit of a segment of the panel.

Photoresistors Ф1 and Ф2 and electroluminescence elements ЭЛ1 and ЭЛ2 are connected in parallel and included consecutively in the circuit of the source of alternating voltage. Element ЭЛ1 is located on the face of the signal panel and glows with excitation. Auxiliary element ЭЛ2 and photoresistor Ф2 are located within the signal panel. Before the inclusion of the segment Ф1 and Ф2 are not illuminated, and therefore a voltage drop across the electroluminescence elements ЭЛ is very insignificant. When Ф1 is illuminated, its resistance decreases and the voltage on elements ЭЛ is increased and they begin to glow. Because of the optical feedback Ф2 is illuminated with light ЭЛ2, and the supply of the activating indicating light can be discontinued.

Besides the described type of panels, which consist of shape segments, abroad matrices with a reference grid are used, and they are more universal since they can be used for the derivation of any graphical information. Figure 8.16 shows the design of the electroluminescent panel with a reference grid.



Metallic conductors (aluminums)
Layer removing the nonlinearity of resistance
Dynode
Phosphorus
Transparent covering
Window
Line of sight

Fig. 8.16. Design of a reference grid of an electroluminescent panel.

Applied to the glass base is a transparent electroconductive covering of zinc oxide, which consists of separate bands (a quantity of bands corresponds to the number of lines in the panel). Applied to this covering is a layer of phosphorus, then a layer of chemically pure metal, which provides the uniformity of the electric field, and a layer which removes the nonlinearity of resistance. Finally, the last layer of aluminum also consists of the separate bands, perpendicular to the bands of the first layer (quantity of bands is equal to the number of columns in the panel). For the luminescence of one point in the matrix, it is necessary to commutate the perpendicular bands, which intersect at this point. The resolution of the coordinate panels is about 1 mm.

8.5. Photochemical and the Electrochemical Devices for the output of Graphical Information

The principle of the operation of photochemical and electrochemical devices is based on the fact that under the action of an electrical pulse or light beam in the emulsion a chemical reaction occurs. Photochemical and electrochemical UVGI most frequently are dynamic indicators of the projector type.

The projector-type dynamic indicators are devices which provide reproduction on the screen of a constantly renewable image.

The projection methods of indication have a high resolving power as a result of the use of an intermediate film data carrier and of a brighter image on the screen because of the capacity of such film carriers to pass intense luminous fluxes.

In this case films of different types, from the ordinary photographic films operating on transllumination to films reflecting light beams, are used.

Devices with the use of the photochromic method of the derivation of graphical information on glass dispositives are most interesting. The image is stored on a special covering of photochromic material with a thickness of about 0.04 mm, in which under the action of a ultraviolet ray a chemical reaction occurs.

The covering is a molecular disperse system of photosensitive dyes with reversible reaction in the continuous (nongranular) material of the covering.

As the separate molecules under the action of the light beam pass over from the transparent to the planted (opaque) state, on the plate there appears a projected image. Since the transparancy of the layer changes at the molecular level, the resolution of the material is very high (more than 1000 lines mm). By means of the photochromic covering it is possible to obtain images with extremely various hues.

The photochromic material in a normal state carries out the visible beams well, remaining in this case transparent. All materials of this type pass over to the painted state with the passage of light close to ultraviolet through them. The rapid return to the transparent state can be realized either by means of heating or by means of illumination by green light.

Furthermore, in the course of time the transparency of the covering increases. In connection with this the observer sees a gradually disappearing trace on the screen. In different photochromic materials the transit time from the opaque state to the transparent fluctuates from fractions of seconds to an hour.

In this case according to the darker margin of the trace the direction of motion of the bobject is determined, and the length of the trace corresponds to its relative speed.

As the light sources for the extraction of the image and characters mercury and xenon tubes are usually used. The image can be recorded on photochromic material and, furthermore, be projected from a slide on a screen. Trajectories of the trace (courses of the motion of objects) are traced in the real time scale by means of movement of the "writing" lenses along the X and Y axes. Each lens focuses an ultraviolet beam on the surface of the photochromic material. It is possible to project several traces on the screen.

Multicolored traces and characters can be obtained by two means: either by using natural colors of photochromic materials or using the mixture of colors. Alphanumeric and other characters can be formed on a photochromic plate by means of masks, i.e., with the formation of the figure light beam up to its input into the collimation lens in the writing device.

The operating principle of projector display units is explained in the example of equipment manufactured by the firm NCR (USA) for the military department (Fig. 8.17).

This system makes it possible to obtain colored images with high resolution on a screen. For the recording of tracks and signs one light source is usually used. This source is used for the illumination of the background grid. The recording can be either semicontinuous (record and storage with photochromic substance) or changeable (projection through a photochromic plate on the screen of indicators or background grids). The tracks of the motion in the real time scale are recorded by means of the transfer of two recording lenses. The lens projects a beam of intense light rays on the photochromic surface. On the plate

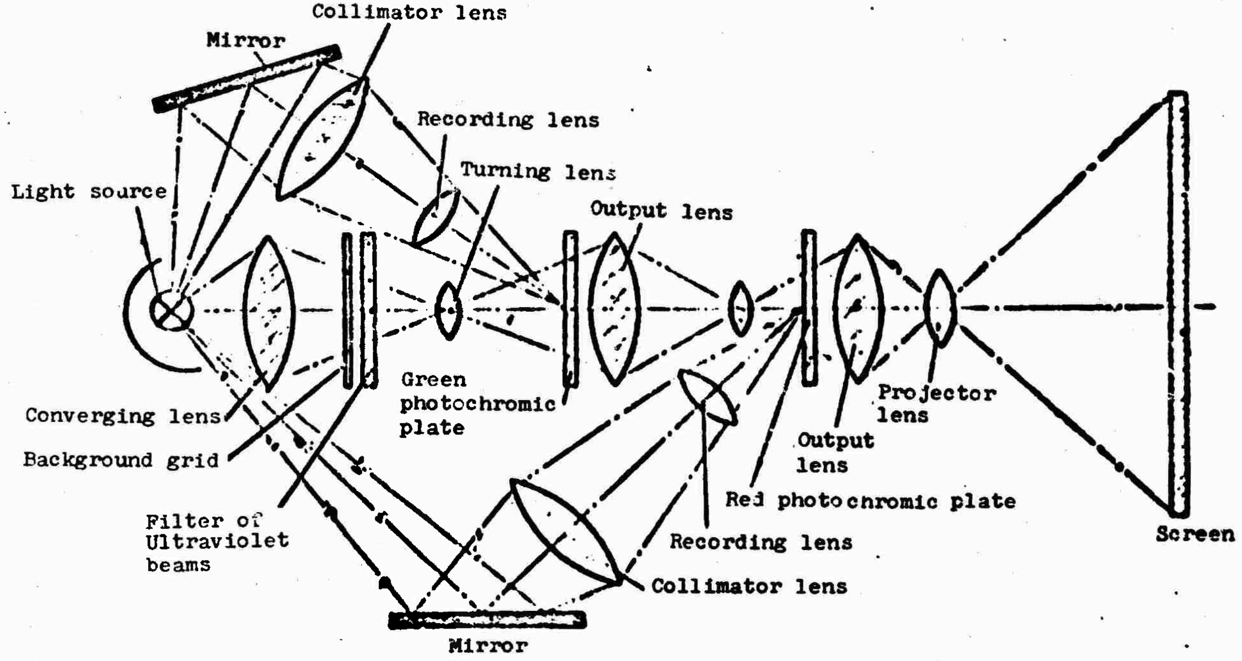


Fig. 8.17. Photochromic color display system of the firm NCR (USA).

the information which is simultaneously projected on a large screen is recorded. Because of the presence of two recording lenses and photochromic plates of green and red color, the information can be reproduced on the screen in two colors - green and red. The scattered light beam passing through the background grid does not affect the photochromic plates.

Above only one of the methods of dynamic indication was shown. In automated control systems other most different methods both in operating principles and in technical fulfilment exist and are used. However, they all are united by one general idea - to obtain a graphic representation of information as far as possible in the real time scale at dimensions of the screens, which allow using this information not for one operator but simultaneously for a large group of those persons who achieve control. Therefore, in many automated control systems the necessary element is the equipment of a large screen.

The dimension of the screen depends on the technical specifications of the equipment and on the methods which are used for obtaining the image. In any case the dimension of the screen is determined basically by the brightness and contrast of the image obtained on it.

Almost all the methods existing abroad of display on a large screen use the method of the direct projection on a large screen of the image obtained on the electronic indicator.

In one of the previous methods was applied the projection directly from the cathode-ray tube screen with dark recording (dark-trace tube). However, this system had considerable shortcomings: little contrast at low repetition frequencies, considerable time for the obliteration of the recorded image and high complexity of the obtaining of an image on a large screen with dimensions of more than 1 m^2.

The photoprojection method has become widespread. It consists of high-speed photography which, accelerated development and immediate projection on a large screen. The entire equipment consists of a camera, equipment for the rapid processing of the motion-picture film and the projection of the processed photograph.

The advantages of the method are the absence of intermediate cathode-ray tube and the possibility of obtaining a clear, bright and high-contrast image on a large (up to 4 m in diameter) screen of constant brightness. The time lag in the transmission of image is the main disadvantage of the method. The time lag of the image on a large screen is determined basically by the time spent on the processing of the photograph.

Such equipment is developed by the firm Canon Instrument (USA). The device 'Repromatik Recorder" makes it possible to reproduce photographs from the Charactron on a large screen 2 seconds after exposure. All the processes in the equipment (photography, processing of film and projection) occur continuously and are controlled by a special synchronizing device. The consumption of the film with the exchange of the frames after 5 seconds for 24 hours of continuous operation consists of a total of 415 m (17,280 frames). The important advantage of the equipment is the possibility of preserving the photographed images as a financial document.

Great possibilities for obtaining a picture of the situation on a large screen are given by the television method. The image from the screen of the Charactron in this case can be photographed by the transmitting television camera the videosignals from which should enter into the television picture tube, which is simultaneously a projector. By means of a special projector the image is transmitted to a large screen. The picture quality on a large screen will depend basically on the technical characteristics of teletron. At the height of the sign on the screen of the Charactron of 3 mm, for its clear reproduction about 10 scanning lines are necessary. If the diameter of the Charactron, for example, is 48 cm, then for the qualitative reproduction of the situation from it a television system with a definition of about 1600 lines is necessary. The obtaining of such a high definition is associated with definite difficulties; however, according to data of the foreign press, such systems are created and used in the Sage system.

For the display of information of a screen with a dimension of 2.4 × 2.4 m, the system SC-2000 is applied, in which the method of xerographic recording is used. In this system (Fig. 8.18) the image from the screen of the Charactron is projected on a selenium plate pre-charged by static electricity. The optical image, in striking on the plate, discharges it on the illuminated sectors. As a result there is created electrostatic relief, which for the development of the image is processed by dyed powder. As a result of such a dry development, whose length does not exceed one second, on the plate there appears a distinct image, which then is projected on the screen. The complete cycle of information processing is 2-5 seconds.

The plate can be used for up to 100,000 times. The image on the screen has high characteristics with respect to brightness, contrast and resolution and is observed without the blackout of the room. The obtaining of a multicolored display is possible.

In certain cases for the representation of a situation of a large screen, display systems using the method of the recording of information of thermoplastic film are used. The film consists of a high-melting base covered with a transparent conductor and a thin layer of low-melting thermoplastic superimposed in the surface of the conductor. The image is recorded by the electron beam on the surface of the thermoplastic film in accordance with

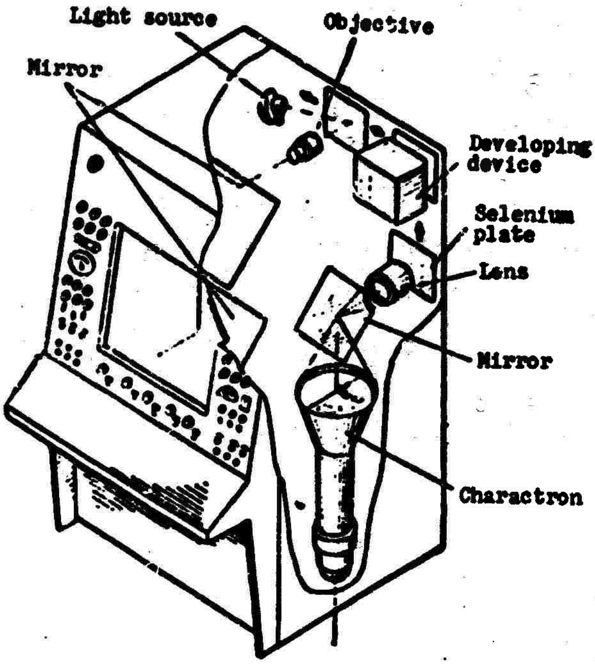


Fig. 8.18. Diagrammatic representation of display system SC-2000.

the incoming information. On the film with respect to the image, i.e., to the charges, the charge pattern is formed. After the heating of the film up to the melting point of the thermoplastic and the subsequent cooling, on the film a pattern is formed which by means of a special optical system is projected on the screen in the form of an image. This method also makes it possible to obtain a colored image.

The recorded information is erased by the neutralization of charges on the film, after which the film is suitable for repeated use.

The system of display which uses the method of thermoplastic recording is very promising. Characteristic of the system are the high density of recording, high resolution, the possibility for rapid reproduction of the recording. The image brightness is sufficient for operation during daylight illumination.

Equipment of the display with the recording of information on thermoplastic film is used in the air defense system 412L (USA), where a large screen with a dimension of 2 × 3 m is used.

8.6. Basic Requirements for the System of the Graphic Display of Information

Devices of the graphic display of information are an integral part of the automated control system.

A block diagram of the standard system of the display of information is represented on Fig. 8.19. The basic elements of any contemporary system of the display of information are: the human operator, the device of the input of information by the operator into the computer, the computer and the information display system. In order to connect the enumerated elements into a single complex, it is necessary that all of them operate within limits of the psychophysiological capabilities of the human operator. In other words, the output data which come from the system must sufficiently simply and accurately answer the question which interest the operator, since not having obtained an answer or not having known how to in time perceive the information, the operator is not able to receive the correct solution. In connection with this, the system of the display of information have a number of requirements, the basic of which are the following:

- high quality of visual perception of the information;
- optimal dimensions of the screen;
- absence of noticeable flickering of the image;

- **optimum illumination of the screen;**

- **possibility of fast extraction of the necessary data by the operator;**

- **optimum cycle of the admission and display of data on the screen;**

- **presence of feedback of the human operator with the display system.**

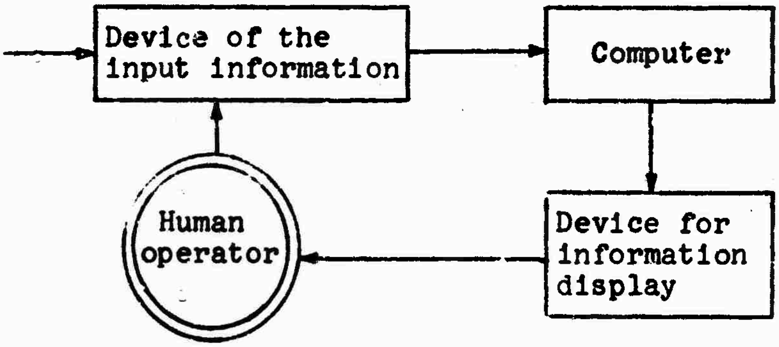

Device of the input information
Computer
Human operator
Device for information display

Fig. 8.19. Block diagram of the standard system of the display of information.

Let us examine these requirements in more detail.

To a considerable degree the quality of the visually received information is determined by the resolution of the display unit. Understood by resolution usually is the degree of discernability of separate parts of the image reproducible on the screen. It is desirable to have such a display unit the resolution of which would correspond to capabilities of the operator's eye. Thus, with normal room illumination the parallel black lines divided by the intervals equal to the width of line are distinguished if the angular dimension of the dividing line lies within limits of 1-1.5'. Thus, the operator's eye can observe separately up to 50 optical lines (100 television lines) per 1°.

Without changing the direction of view, the operator can see a two-dimensional image the angular dimension of which is 50°. This makes it possible to distinguish on the indicator up to 2500 optical lines. In this case it is necessary to keep in mind that the resolution is changed depending on the contrast and image brightness, the presence of noise the mixing interaction of the auxiliary image, the time of the exposure of the image, and the surrounding situation.

In speaking about the dimensions of the screens, it is necessary to remember that an increase in the screen does not always lead to an improvement in the operating conditions. Thus, for instance, in system where there are used large indicators of collective use the work of the operator is made difficult. Several operators work behind one screen in such systems. Each operator, in solving his problem, selects on the screen only information of interest to him. In this case the data which are of interest for only one operator prevent the work of other operators, impeding their selection of the necessary information.

The flickering of the image leads to the premature fatigue of the operators. This harmful phenomenon is illuminated by red reducing the frequency of the exchange of frames of the image up to 26-40 per second.

The selection of illumination of the screen of the display unit is a rather complex problem. With great illumination of the screen the operator distinguishes poorly the elements of the reproducible image. In this case the working conditions of the operator deteriorate in connection with a decrease in the contrast. External illumination creates one additional difficulty. Being reflected from the surface of the screen, light from external sources can sharply lower the image contrast.

A very important requirement given to the system of the display is the capability of rapid distinguishing by operator of the necessary data with the least amount of action and computations.

In developing systems of the display of information, it is necessary to lead, in accordance with the carrying capacity of the human operator, the quantity of obtained information to the storage and processing of data. On the average the limit of the retention of information in the immediate-access memory of man does not exceed 3 seconds. After a somewhat long time the probability of the retention of information in the memory of man sharply drops, especially in such a case when after other messages have entered (after 18 seconds the probability of the memorization of even of simple and familiar information will be only 10%). Therefore, one of the basic requirements for display systems is the limitation of the cycle of the admission and display of data by a time not exceeding 3 seconds.

The assurance of the operator in the reliable operation of the display system is one of the important factors. Such an assurance can be provided by means of the introduction of feedback into the system, which is especially necessary when the entering message is quite large.

CHAPTER 9

DEVELOPMENT OF THE SOLUTION FOR COMBAT OPERATIONS OF ACTIVE MEANS

9.1. Basic Problems Solved by Computers in the Control of Air Defense Weapons

The outcome of combat operations under contemporary conditions in many respects is determined by the flexibility and operational efficiency of the system of control of the forces. Combat operations of air defense weapons are so transient that the leader is not able in proper time to react to changes in the situation and make the correct decisions. At the present time the problem of the delivery of the optimum solution is placed on the computer, which implements that algorithm which was previously developed and "placed" into the computer.

For the automation of the solution of problems which appear during the planning and preparation of combat operations, their formalization is necessary, i.e., such a special (mathematical and logical) formulation of operative problems which makes it possible to create algorithms for the operation of computer and, therefore, makes their use possible. During the creation of the algorithms of combat operations, in a number of cases it is important to formulate correctly the "logic of battle" and develop criteria of the effectiveness of the problem being solved. Such algorithms are created usually by the combined efforts of the officer-tacticians, mathematicians and programming engineers.

The basic problems being solved on the computer of the automated control systems are:

- the processing of incoming information and the solution of the information and logic problems and the coding and the decoding of information;
- the evaluation of the effectiveness of the armament utilized and the selection of the most effective combat means under specific conditions;
- the solution of the problem of target assignment, i.e., the rational application of active means (interceptors, guided missiles, etc.) at targets;
- the development of calculation data necessary to the leader for making a decision;
- the data processing connected with logistic support, storage, account and distribution of materials;
- calculations connected with combat training, and so on.

Because of the close connection of mathematical methods of the investigation of combat operations with computer technology, the latter acquires an ever increasing value. One should, however, consider that not any, even the most perfected, computer is capable of completely replacing man, to whom the creative and intellectual qualities and the latitude of the envelopment of the phenomena are characteristic. Let us emphasize that most frequently the computers which operate in forces produce not the solution but only the principle, which helps the leader (commander) to make the correct decision. In other cases the computer only implements that decision which was prepared by leader for an analogous case previously in a calmer situation. It is necessary to bear in mind that any mathematical formulation of combat operations is only a scheme, and it does not reflect the true relationship of the opposing sides. There are many factors (political, moral, etc.) which at the present time cannot be described mathematically but which can substantially affect the course and outcome of the combat operations.

In connection with this the leader should not "follow" blindly behind the computer. He takes the final solution on the basis of the comprehensive examination of the situation and analysis of quantitative and qualitative criteria, taking into account factors not yielding to mathematical evaluation. Thus, the creative process of making a decision for carrying out combat operations was and remains that of man.

The use of computational and data processing computers frees the staff workers and leader of the need to store in the memory an enormous quantity of secondary information. The possibility of creating an effective control system, the saving of time in the solution to problems connected with the planning and conducting of combat operations, the high accuracy and reliability of the solution of problems, the automatic account of the rapidly changing factors, the freeing of people from the tedious work, etc., are some of the advantages which use of the computer gives for the solution of problems of the control of air defense combat operations.

9.2. Basic Principles of Target Assignment

Target assignment is called the operation consisting in the assignment of a definite target to the selected fire means. If there are several targets which must be subjected to fire interference, when in the solution of the target-assignment problem it should be accurately shown which means (antiaircraft guided missiles, fighters or antiaircraft guns) should be used, in what quantity, and when should they be directed to each of the targets subjected to bombardment.

Under conditions of the repulsing of great air raid, the combat situation can be so complex, and the number of possible versions of target assignment so great that making a decision without special estimates proves to be impossible. In order that the target-assignment problem could be transmitted by the computer, the algorithm of the solution of this problem (algorithm of target assignment) should be created.

The problem of target assignment in full volume is very complex and requires an account of many factors, such as, for example:

- the deployment of distributed weapons and their combat readiness;

- the transmission capacity of the guidance channels and communication channels;

- the depth of the zone scanned by radar sets;

- the possibility of the use by the attacking side of a maneuver, interference, etc.

All these circumstances should in one way or another be considered during the composition of the algorithm of target assignment, which is preliminarily played on a number of the models of battle in order to select the most adequate one. As usually occurs in the solution to complex problems of operational research, here it is not possible to be limited to an evaluation according to one criterion (index) of effectiveness, but one should search for a compromise solution which is satisfactory according to a whole series of criteria.

Implied under the effectiveness of the algorithm is the degree of its adaptation to the fulfillment of the problems placed before it. The better the algorithm is formed, the more effective it is.

However, in order to judge the effectiveness of the algorithm of target assignment, it is necessary to have a certain numerical criterion, which can be called the index of effectiveness.

Used as an index of effectiveness M is usually either a probability of any event or mean value (mathematical expectation) of a certain random value. For example, the indexes of effectiveness can be the destruction probability of a target, the middle area of destruction on a certain object, and so on.

The specific form of the index of effectiveness is selected depending on the problem being solved. By examining different problems from the viewpoint of their purposeful directivity, it is possible to note two limiting cases. In the first case the operation is fulfilled for the purpose of the achievement of a completely definite result (it is necessary to hit one aircraft or all aircraft in the attacking group, etc.). This result can or cannot be achieved, and no intermediate results are examined. In other words, the success of the operation can be evaluated according to the scheme "yes"-"no" ("all" or "nothing").

In such case the natural index is the probability of the achievement of the desired result or the probability of the fulfillment of the combat mission.

If we designate the event which consists in the fact that the combat mission is fulfilled by γ, then the index of effectiveness $M = P(\gamma)$ is a probability of the event.

In the second case the mission of the operation is the applying to the enemy of the maximal possible damage ("the more, the better"). In such cases the natural index of effectiveness will be the mean value (or mathematical exception) of the damage applied to the enemy:

$$M = m[Y], \qquad (9.1)$$

where the random variable Y is the damage caused; m is the index of mathematical expectation.

Depending on the targets and form of operations, the value Y can have one or another sense (area, volume, time, number of destroyed targets, etc.) and, correspondingly, can be expressed in those or different units.

Frequently as an index effectivenesses we take not simply the average damage but the average relative damage (for example, the average portion of the destroyed targets in the composition of group or the average portion of the destroyed area).

As an illustration let us examine two examples of the selection of the index of effectiveness.

1. A grouping of the antiaircraft guided missiles which defends a zone (let us say near the border) conducts fire on a group of attacking bombers. The mission of the antiaircraft guided missile [SAM] (ЗУР) is to knock as large a number of the

opponent's aircraft as possible. The index of effectiveness in this case is the average number of destroyed aircraft or the average portion of destroyed aircraft in the composition of the raid.

2. The grouping of the antiaircraft guided missiles, which defends one object, conducts fire on a group of aircraft attempting to break through to the object. Each of the aircraft can be the carrier of a powerful weapon of destruction, and therefore the penetration of at least one aircraft is practically equivalent to the obliteration of the object. The defense's mission is not to let one of the opponent's aircraft penetrate. Here the index of effectiveness is the probability that not one aircraft will penetrate to the object.

In these examples the index of effectiveness should be greatest. In general this is not compulsory. It is possible to use such indexes which, the contrary, it is desirable to have smallest. Thus, it is possible to take as an index of effectiveness the minimum of the damage which the opponent's aircraft can inflict.

Since the final result of combat operations of the air defense system is considered as the obliteration of the air attack weapons or the nondestruction of the defended objects, as an index of effectiveness it is advantageous to take either the maximum of the mathematical expectation of opponent's destroyed aircraft or the minimum of the mathematical expectation of the damage. Moreover, depending on the missions fulfilled by the air defense system one or another criterion can be applied. For the small units whose basic mission is the obliteration of the air attack weapons of the enemy the first criterion is more acceptable; for small units which defend a specific object the second criterion is.

To understand the principles of the target assignment, it is useful to examine several of the simplest schematized problems of target-assignment which explain [illegible] methods of the decision making and show what advantage in [illegible] iveness is given by the correct solution to problem.

Let us assume that before the grouping of antiaircraft missiles which armed with n complexes of SAM there is the problem of the obliteration of the dispersed group consisting of N targets. Let us assume, for simplicity, that each complex of the SMA can launch only one missile at any target in the group. The destruction probability by the i-th complex of the j-th target is assigned equal to P_{ij}. Values of P_{ij} are recorded in Table 9.1.

Table 9.1.

Number of complex	Number of target			
	1	2	. . .	N
1	P_{11}	P_{12}	. . .	P_{1N}
2	P_{21}	P_{22}	. . .	P_{2N}
. . .	. . .	. . .	. . .	. . .
. . .	. . .	. . .	. . .	. . .
. . .	. . .	. . .	. . .	. . .
n	P_{n1}	P_{n2}		P_{nN}

It is necessary to find the optimum (best) target assignment, having assigned to each complex of the SAM a definite target which the latter must destroy (it is possible that the same target will be fired at by several complexes).

The stated problem is called the target-assignment mission n × N. For its solution let us first select the index of effectiveness.

If the grouping of the SAM is intended for the obliteration of the air attack weapons of the enemy, then as an index of effectiveness let us take the maximum of the mathematical exception of the destroyed aircraft:

$$M_n = m[X_n].$$

where the random variable X_n is the total number of destroyed targets.

Let us represent the total number of destroyed targets X_n in the form of the sum of N random variables:

$$X_n = X_1 + X_2 + X_3 + \ldots + X_N = \sum_{i=1}^{i=N} X_i.$$

Each i-th target corresponds to its random value defined as:

$$\begin{cases} X_i = 1, \text{ if the target is hit;} \\ X_i = 0, \text{ if the target is not hit.} \end{cases}$$

It is not difficult to ascertain that the total number of destroyed targets X_n is equal to the sum all values X_i. According to the theorem of the addition of the mathematical expectations

$$m[X_n] = m[X_1] + m[X_2] + \ldots + m[X_N] = \sum_{i=1}^{i=N} m[X_i]. \qquad (9.2)$$

Let us designate by P_i the destruction probability of the i-th target as a result of the bombardment of it by missiles of the complex. Then by definition of the mathematical expectation

$$m[X_i] = P_i \cdot 1 + (1 - P_i) \cdot 0 = P_i.$$

Substituting into (9.2), we obtain

$$m\{X_a\} = P_1 + P_2 + \ldots + P_N = \sum_{i=1}^{i=N} P_i$$

or finally

$$M_a = \sum_{i=1}^{i=N} P_i$$

i.e., the average number of targets hit in the composition of the group is equal to the sum of the destruction probabilities of all the separate targets.

Thus, with the target assignment according to mathematical expectation it is necessary to distribute complexes on the targets in order that the sum of the destruction probabilities reach a maximum. The simplest method of the solution of the indicated problem consists in the sorting of all possible variants of the distribution of complexes on targets and in the selection of those for which the sum of the probabilities reaches a maximum.

The approximate target-assignment problem of two aircraft by two antiaircarft missile complexes (the so-called 2 × 2 problem) is represented in the form of Table 9.2.

Table 9.2.

(1) Номер i комплекса	(2) Номер j цели	
	1	2
1	$P_{11} = 0.8$	$P_{12} = 0.6$
2	$P_{21} = 0.7$	$P_{22} = 0.1$

KEY: (1) Number i of the complex; (2) Number j of the target.

The values of destruction probability by complexes of SAM of targets are given conditionally in the table as an example.

At first glance it can be shown that the best variant of target assignment consists in the assignment of the first target to the first complex ($P_{11} = 0.8$) and the second target to the second complex ($P_{22} = 0.1$). However, it is easy to be convinced of the fact that this is erroneous, having sorted out other possible variants (a total of 4 of them).

Let us write each variant in the form of a column (Table 9.3), where i is the number of the complex, and j is the number of the target.

Table 9.3.

Вариант 1 (1)		Вариант 2 (1)		Вариант 3 (1)		Вариант 4 (1)	
i	j	i	j	i	j	i	j
1	1	1	1	1	2	1	2
2	1	2	2	2	2	2	1

KEY: (1) Variant.

Let us determine for each of the variants the mathematical expectation M_n of the number of targets hit.

In the first variant both complexes fire at the first target: the probability of its destruction is $P_1^1 = 1 - (1 - P_{11})(1 - P_{21})$.

Substituting here values of the destruction probabilities of the first target by each of the complexes (Table 9.2), we will obtain $P_1^1 = 1 - (1 - 0.8)(1 - 0.7) = 0.94$.

Since firing is not conducted at the second target, the probability of its destruction is $P_2^1 = 0$.
Hence

$$M_n^1 = P_1^1 + P_2^1 = 0,94$$

For the second variant

$$M_n^2 - P_1^2 + P_2^2 - 0,8 + 0,1 = 0,9$$

For the third variant

$$M_n^3 - P_1^3 + P_2^3 = 1 - (1 - 0,6)(1 - 0,1) = 0,64$$

For the fourth variant

$$M_n^4 = P_1^4 + P_2^4 = 0,6 + 0.7 = 1,3$$

Thus, the fourth variant is the optimum, i.e., it is necessary to direct the first complex to the second target and the second complex to the first target.

Naturally, under actual conditions of the conducting of combat operations, the quantity of air attack weapons of the enemy and, correspondingly, the number of antiaircraft missile complexes can be considerably greater and be changed within sufficiently wide limits. Then the simple sorting of all possible variants becomes complex even for the computer. There appears the need for the solution of the target-assignment problem without the sorting of all variants. The most optimum methods which make it possible to solve the target-assignment problem by such means are methods of linear programming.

9.3. Optimization of the Index of Effectiveness by Means of Methods of Linear Programming

Linear programming is the mathematical method making it possible to find the most advantageous formalized solutions not only of the target-assignment problem but also a number of other tactical and operative problems. For example, for the antiaircraft defense of several objects, it is possible to find the most advantageous distribution of combat means, ammunition, critical sectors, etc.

For this purpose one of the values (let us assume the number of hits by the SAM into the enemy's aircraft) is expressed in the form of a linear function of the number of variables whose distribution is determined. The indicated function can be expressed in the following form:

$$M = C_1X_1 + C_2X_2 + \ldots + C_nX_n \qquad (9.3)$$

where M is the final result of the distribution of resources (fighter aircraft, SAM, etc.); X_1, X_2, ..., X_n - the variables on that or a different value of which the result of the distribution depends; C_1, C_2, ..., C_n - the parameters which determine the utilization factor of the variables being investigated in this or a different method.

For the solution of equation (9.3), reduction to the maximum or minimum of value M, it is necessary to compose and solve in conjunction with equation (9.3) a number of linear equations or the inequalities which describe the limitations which are imposed by those or different conditions of the solution of problem, depending on the situation.

These limiting inequalities or equations usually take the following form:

$$\begin{aligned} &A_{11}X_1 + A_{12}X_2 + \ldots + A_{1n}X_n > b_1; \\ &A_{21}X_1 + A_{22}X_2 + \ldots + A_{2n}X_n > b_2; \\ &\ldots\ldots\ldots\ldots\ldots\ldots \\ &A_{m1}X_1 + A_{m2}X_2 + \ldots + A_{mn}X_n > b_m; \end{aligned} \qquad (9.4)$$

in this case $X_1 \geq 0$, $X_2 \geq 0$, ..., $X_n \geq 0$.

Values X_1, X_2, ..., X_n have in expression (9.4) the same value as in expression (9.3). Values A_{ij} and b_j are the coefficients determined from conditions of the solution of the problem.

Let us examine the possibility of solving the target-assignment problem by methods of linear programming in the example of the following tactical situation. There are two objects I and II defended two antiaircraft complexes No. 1 and No. 2. The objects are raided by two of the enemy's aircraft (targets 1 and 2) with the different striking capabilities.

If the first target penetrates to object I, then it will inflict damage of six arbitrary units, and if to object II - then four conditional units. The possibilities for target 2 is two times less, i.e., are correspondingly equal to three and two conditional units.

If any target bracketed by complex No. 1 attacks object 1, then it is destroyed with the probability $P_1 = 1$. If this aircraft strikes object II, then due to an increase in the firing distance it will be destroyed with the probability $P_2 = 0.8$. In this case it is assumed that the complex of the SAM does not manage to fire at another aircraft. The situation is similar also with complex No. 2. Thus, each complex can simultaneously fire at only one target.

It is necessary to determine how advantageous it is to use the complexes for firing at the targets in order that the damage inflicted by the enemy would be minimum.

For the one who is being defended there are four possible methods (rules) of the use of complexes called strategies:

- bracket by complex No. 1 (and escort) target 1, and by complex No. 2 - target 2;

- bracket target 2 by the complex No. 1, and by complex No. 2 - target 1;

- bracket target 1 by both complexes;

- bracket target 2 by both complexes.

For the one who is attacking there are also four strategies:

- target 1 strikes object I, target 2 strikes object II;

- target 1 strikes object II, target 2 strikes object I;

- both targets strike object I;

- both targets strike object II.

Let us designate by X_{ij} and Y_{ij}, respectively, the unknown strategies of the enemy and the defender, where

$$X_{ij} = \begin{cases} 1, \text{ if the i-th aircraft attacks the j-th object;} \\ 0 - \text{in the opposite case;} \end{cases}$$

$$Y_{kl} = \begin{cases} 1, \text{ if the k-th complex bracket the } l\text{-th target;} \\ 0 - \text{in the opposite case.} \end{cases}$$

Let us produce calculation of the mathematical expectation of the inflicted damage according to formula known in the theory of games [25]:

$$C_{mn} = 6X_{11}Y_{12}(0{,}2Y_{21} + Y_{22}) + 3X_{12}Y_{11}(Y_{21} + 0{,}2Y_{22}) + \\ + 4X_{21}Y_{22}(0{,}2Y_{11} + Y_{12}) + 2X_{22}Y_{21}(Y_{11} + 0{,}2Y_{12}). \qquad (9.5)$$

For example, for the first strategy of the defender and second strategy of the attacker we have:

$$X_{11} = 0;\ X_{22} = 0;\ X_{12} = 1;\ X_{21} = 1;$$
$$Y_{11} = 1;\ Y_{22} = 1;\ Y_{12} = 0;\ Y_{21} = 0;$$

and on the basis of formula (9.5) we obtain $C_{12} = 0.6 + 0.8 = 1.4$.

If we sort out all the possible combinations of strategies of the opponent and defender, then the value of mathematical expectation of the damage inflicted by the opponent for each such combination can be presented in the form of a table called the matrix of damages.

The matrix of damages calculated from formula (9.5) is given in Table 9.4.

Table 9.4.

		No. of strategy of the defender			
		1	2	3	4
Number of opponent's strategy	1	0	1,6	2.0	6,0
	2	1,4	0	3,0	4,0
	3	0,6	1,2	3,0	6,0
	4	0,8	0,4	2,0	4,0

Let us note that strategies 3 and 4 of the defender with any method of actions of the attacker for it are less advantageous than strategies 1 and 2, and therefore (columns 3 and 4 of Table 9.4) they can be excluded from examination.

In the terminology of the theory of games the target-assignment problem being solved is called a rectangular or matrix game. It consists of the following.

The first player selects a certain positive integer number $m(1 \leq m \leq M)$, and the second player, without knowing about what selection the first made, selects a certain number $n(1 \leq n \leq N)$.

Then these two numbers are compared, and one of the players (for example, the first) pays another sum $A_{m,n}$, which depends on the selections made and is determined by rules of the game. The set of numbers $A_{m,n}$ forms a rectangular table called the payoff matrix. If the second player makes a selection, by knowing the selection of the first, then such a game is called the simplest positional game.

The selection of a definite number by a participant of the game is called pure strategy of the player. The theory of games indicates the method of the conducting of games of this type. The first player should select the strategy m with a certain probability P_m $(m = 1, 2, 3, \ldots, M)$, whereupon $P_1 + P_2 + \ldots + P_M = 1$. The set of these probabilities is designated by vector $M = (P_1, P_2, \ldots, P_M)$ and is called the mixed strategy of the first player.

Analogously the second player selects the strategy n with probability P_n $(n = 1, 2, \ldots, N)$. The corresponding vector is designated $N = P_1, P_2, \ldots, P_N$.

In this case the mathematical expectation of the payment

$$M(m, n) = \sum_{m=1}^{M}\sum_{n=1}^{N} A_{mn}P_mP_n$$

Value P_m is selected in such a way that each of the conceivable mixed strategies of the first player m is contrasted to the most advisable (with respect to it) mixed strategy of the second player, giving to him the payoff $M_1(m) = \max M_1(m, n)$, and then the first player is stopped on the mixed strategy, minimizing function $M_1(m)$. As a result of such an approach to the selection of mixed strategy the first player's payment is

$$M_1 = \min_m M(m) = \min_m \max_n M(m, n) = \min \max M(m, n).$$

Value n is selected by the second player according to analogous considerations, and his payoff in this case is $M_2 = \max \min M(m, n)$.

It is mathematically proved that values M_1 and M_2 coincide. Furthermore, if one of the sides is held by the optimum mixed strategy selected by it, and the other deviates from it, then from this the deviating side can only lose, but in no case does it increase its payoff.

In approaching to the solution of the rectangular game, by the assigned matrix of damages for the defender let us check whether or not it is possible to obtain the solution to the game in pure strategies. By applying the first pure strategy, the defender can sustain a damage of 1.4 arbitrary units; by applying the second strategy, the defender can sustain a damage of 1.6 arbitrary units.

Thus,

$$M_1 = \min_m (\max_n A_{m,n}) = \min (1,4;\ 1,6) = 1,4.$$

By conducting analogous calculations for the rival, we obtain

$$M_2 = \max_{s}(\min A_{m,s}) = \max(0;\ 0;\ 0{,}6;\ 0{,}4) = 0{,}6.$$

The noncoincidence of values M_1 and M_2 indicates the fact that the solution of game should be sought in mixed strategies. For finding solutions of rectangular games there are many methods. Let us solve the target-assignment problem by means of the analytical method, having reduced it to the problem of linear programming.

Let us designate the unknown probabilities of the use of strategies by the defender by P_1 and P_2. The thus far unknown greatest damage inflicted by the opponent is designated by M.

If we assume that the attacking side, having deviated from the optimum mixed strategy, will apply its first strategy, then on the average it will inflict on the defended objects the damage $C_{11}P_1 + C_{11I}P_2 = 0P_1 + 1.6\ P_2 = 1.6\ P_2$, which cannot be more than the maximum damage M inflicted by the opponent when using the optimum mixed strategy, i.e.,

$$1{,}6P_2 < M. \qquad (9.6)$$

Similarly, for the second, third and fourth pure strategies of the attacking side:

$$1{,}4P_1 < M; \qquad (9.7)$$

$$0{,}6P_1 + 1{,}2P_2 < M; \qquad (9.8)$$

$$0{,}8P_1 + 0{,}4P_2 < M. \qquad (9.9)$$

In this case it is evident that $P_1 + P_2 = 1.$ (9.10)

Let us divide inequalities (9.6)-(9.9) and also equality (9.10) by M, and let us introduce the new notations:

$$X_1 = \frac{P_1}{M};\ X_2 = \frac{P_2}{M};\ \frac{1}{M} = L \qquad (9.11)$$

Then the total value of the expected damage

$$L = X_1 + X_2 \qquad (9.12)$$

in this case X_1 and X_2 should satisfy the following conditions:

$$\begin{gathered} 1{,}4X_1 < 1; \\ 1{,}6X_1 < 1; \\ 0{,}6X_1 + 1{,}2X_2 < 1; \\ 0{,}8X_1 + 0{,}4X_2 < 1, \end{gathered} \qquad (9.13)$$

which correspond to requirements of strategies of the defender and attacker.

In order to convert the system of inequalities (9.13) into an equality in accordance with requirements for conditions of linear programming, it is possible to introduce additional values X_3, X_4, X_5 and X_6 (frequencies of the selection of strategies not being utilized by the attacker and defender but also not affecting the value of the expected damage) and rewrite conditions (9.13) in the following form:

$$\begin{gathered} 1{,}4X_1 + X_4 = 1; \\ 1{,}6X_2 + X_3 = 1; \\ 0{,}6X_1 + 1{,}2X_2 + X_5 = 1; \\ 0{,}8X_1 + 0{,}4X_2 + X_6 = 1. \end{gathered} \qquad (9.14)$$

By selecting the probabilities (frequencies) P_1 and P_2 and the values X_1 and X_2 connected with them, the defender attempts to make the magnitude of the expected damage M as less as possible, and, consequently, value L as large as possible.

By solving by means of methods of linear programming the problem on the maximization of value L with the obtained limitations, it is possible to obtain such frequencies of the selection of strategies by the defender which allow it to reduce the expected damage to a minimum.

The basic method of the solution of such problems is the so-called simplex method (the method of sequential improvement in the plan). The solution to the problem by means of the simplex method is divided into two stages. In the first stage we find any (even if very unsuccessful) solution satisfying the sets of linear limitations, or we are convinced of the fact that such a solution does not exist. This stage is called searching for the initial plan. In the second stage the sequential improvement of the initial plan is conducted according to definite rules until further improvement becomes impossible. Let us agree to call the variables which enter into the plan the basic and the variables not entering into the plan nonbasic.

Computations are carried out from the following algorithm:

a) such an initial plan is sought in order that the basic variables would be expressed as nonbasic variables;

b) the value of the function L being minimized is expressed in terms of nonbasic variables;

c) selected is the one of the nonbasic variables the introduction of which into plan is capable of improving value L;

d) determined is which of the basic variables should be excluded from the plan and made nonbasic;

e) the variable newly introduced into plan is expressed as the variable derived from the plan and other nonbasic variables;

f) all the remaining basic variables and the values of the minimized function L are expressed as new nonbasic variables;

g) operations of items b-f are repeated.

If at any stage it appears that the introduction into the plan of any of the nonbasic variables is not capable to decrease (increase) the value M, then the latter plan proves to be the best, and the value L corresponding to it smallest (largest).

It is proved that if we follow this algorithm, then through the final and usually small number of steps the solution to the problem is reduced to the optimum plan.

For an understanding of the content of the simplex method, let us examine in more detail the solution of the target-assignment problem formulated above which is necessary for obtaining the least possible damage.

In the problem it was necessary to maximize expression (9.12) with limitations (9.13). In accordance with the planned algorithm let us find the initial plan. It is necessary to note that in general the finding of the initial plan is no less difficult a problem than the problem of its improvement.

Let us solve system (9.13) relative to the newly introduced variables:

$$X_3 = 1 - 1{,}6X_2; \qquad (9.15)$$

$$X_4 = 1 - 1{,}4X_1; \qquad (9.16)$$

$$X_5 = 1 - 0{,}6X_1 - 1{,}2X_2; \qquad (9.17)$$

$$X_6 = 1 - 0{,}8X_1 - 0{,}4X_2. \qquad (9.18)$$

As the basic variables let us take X_3, X_4, X_5 and X_6 and as nonbasic variables - X_1 and X_2. In accordance with item b of the algorithm let us express function L by nonbasic variables.

Substituting $X_1 = X_2 = 0$ into expressions (9.15-9.18), we obatin

$$X_3 = X_4 = X_5 = X_6 = 1;$$
$$L = 0.$$

Let us discuss the variable X_4 determined from expression (9.16): $X_4 = 1 - 1.4X_1$. It indicates that with the introduction of value X_1 into the plan, the variable X_4 should be excluded from the plan and made a new nonbasic variable.

In accordance with item e and using expression (9.16), we obtain $1 - 1.4X_1 = 0$, whence

$$X_1 = 0.714. \quad (9.19)$$

Substituting (9.19) into (9.17) and (9.18), we have $X_5 = 0.572$, $X_6 = 0.428$.

From expression (9.15) the value $X_3 = 1$ is determined.

The obtained plan appears in the following manner: $X_2 = X_4 = 0$; $X_1 = 0.714$; $X_3 = 1$; $X_5 = 0.572$, $X_6 = 0.428$, and in accordance with (9.12) $L = 0.714$.

In continuing in a similar way, let us introduce variable X_2 into the plan. Having made X_2 a new basic variable and having recorded the fact of an increase in value L with the introduction into the plan for the variable X_2, let us exclude one of the variables X_3, X_5 or X_6 from the plan. It is not difficult to show that the derivation from the plan of variable X_3 or X_6 will lead to a disturbance of limitations imposed on (9.12) according to the condition of the problem.

Actually, if $X_3 = 0$, then, as follows from (9.15), $X_2 = 0.625$. By substituting values $X_1 = 0.714$ and $X_2 = 0.625$ into expression (9.25), we obtain $X_5 = 1 - 0.428 - 0.750 = -0.178$, which is inadmissible according to the condition of the problem. The same is obtained if $X_6 = 0$. In this case from (9.18)

$$X_2 = \frac{1 - 0.2X_1}{0.8} = 1.07.$$

Substituting this value into (9.15), we obtain $X_3 = 1 - 1.6X_2 = -0.712$

Now the target is reached. As an initial plan it is possible to take $X_1 = X_2 = 0$, $X_3 = X_4 = X_5 = X_6 = 1$, and value L, opposite to the value of the expected damage, will in this case be equal to 0.

Let us try to increase L. From expression (9.12) it is apparent that the introduction of X_1 into the plan is capable of increasing L. Let us see in what dimensions it is possible to allow the introduction into the plan of variable X_1.

The equality $X_4 = 1 = 1.4X_1$ shows that X_1 can be introduced into the plan at a value not exceeding 0.714 (otherwise X_4 will take a negative value).

With the introduction of variable X_1 into the plan, it should become the basic variable, and one of the variables X_3, X_4, X_5, and X_6 should be excluded from the plan, i.e., should become a new nonbasic variable (variable X_2 remains nonbasic). In accordance with item d of the algorithm, let us determine this new nonbasic variable. From equality (9.15) it follows that X_3 cannot be excluded from the plan, since otherwise this equality will be destroyed.

It is also not possible to make the variable X_5 nonbasic, since, as follows from expression (9.17), $X_5 = 1 - 0.6X_1 - 1.2X_2$, and in the case $X_5 = 0$ value X_1 becomes inadmissibly large:

$$X_1 = \frac{1}{0.8} - \frac{1}{0.8} X_6 = \frac{1}{0.8} = 1.25$$

Let us recall that value X_1 cannot exceed the value equal to 0.714. For this reason the variable X_6 also cannot be derived from the plan, since with equality $X_6 = 0$, $X_1 = 1.25$.

Thus, with the introduction of X_2 into the plan, variable X_5 should be excluded from the plan and made nonbasic variable. Let us express the variable X_2 being introduced into the plan by a variable X_5 derived from the plan:

$$X_2 = 0.83 - 0.5X_1 - X_5 \tag{9.20}$$

Since $X_1 = 0.714$ (since X_4 is also a nonbasic variable), X_2 in accordance with (9.20) will be defined as $X_2 = 0.476$. Substituting X_1 and X_2 into (9.15) and (9.18), we will obtain X_3 amd X_6: $X_3 = 0.236$, $X_6 = 0.238$.

Let us determine value L: $L = X_1 + X_2 = 1.190$. The obtained plan $X_4 = X_5 = 0$, $X_1 = 0.714$; $X_2 = 0.476$; $X_3 = 0.236$; $X_6 = 0.238$; $L = 1.190$.

The circumstance that the introduction of X_4 or X_5 into the plan does not increase values L indicates that the obtained plan is the best.

Hence, by using expression (9.11) we will obtain $P_1 = 0.6$; $P_2 = 0.4$; $M = 0.84$.

The obtained solution is interpreted in the following manner. The defender should use either the first pure strategy (complex No. 1 fires at the target 1, complex No. 2 - target 2), or the second (complex No. 1 fires at target 2, complex No. 2 - target 1). In this case the selection of the strategy should be determined with probabilities $P_1 = 0.6$ and $P_2 = 0.4$, respectively.

The positive feature of the simplex method is its generality: it is applicable for the solution of any problem of linear programming. Its basic shortcoming is its unwieldiness. In the solution to problems with a large quantity of variables, a very large amount of time is expended for their solution, which frequently with the conducting of combat operations is inadmissible. In connection with this, such modifications of the simplex method are used which are more simply realized in the form of algorithms.

Let us examine the solution to the target-assignment problem by means of one of the modifications of the simplex method - Hungarian method (thus named in memory of the Hungarian mathematician Koenig).

9.4. Solution of Target-Assignment Problems by the Hungarian Method

For the solution of target-assignment problems the matrix of damages is compiled:

$$C = \begin{vmatrix} C_{11}C_{12}C_{13}\ldots C_{1n}; \\ C_{21}C_{22}C_{23}\ldots C_{2n}; \\ \ldots C_{ij}\ldots; \\ C_{m1}C_{m2}C_{m3}\ldots C_{mn} \end{vmatrix} \qquad (9.21)$$

where C_{ij} is the damage inflicted by the attacker on the defender with the application of the i-th active means on the j-th target.

It is necessary to distribute thus the active means on the targets in order that the expected total damage would be minimum. In principle the optimum solution of the problem can be obtained if we sort out all the possible solutions and select the best one. The number of such solutions depends on the number of lines and columns in the matrix of damages.

For example, for the matrix with 5 lines and 5 columns in all there can be $1 \cdot 2 \cdot 3 \cdot 4 \cdot 5 = 120$ solutions. With 6 lines and 6 columns, solutions for comparative evaluations are already $1 \cdot 2 \cdot 3 \cdot 4 \cdot 5 \cdot 6 = 720$.

If the matrix has 20 lines and 20 columns (20 aircraft of the attacker against 20 active means of the defender), then the number of all solutions will reach $2 \cdot 4 \cdot 10^{17}$. If for the composition of one solution the computer expends 1 μs, then 7.74 thousand years will be required for the finishing of the solution to the problem to the very end. Such problems in which we are speaking of more than about six distributions, as a rule, are not solved by the enumeration of all solutions (with the exception of separate, very special cases). At the same time the successful solution of the target-assignment problem can be achieved if we use the Hungarian method, the essence of which is explained in the following example.

Five aircraft produce a raid on objects of the area which are covered by five active means (uniform or different). The matrix of the damages selected for an example randomly in this case takes the following form (Table 9.5).

Table 9.5.

Номер активного средства, j (1) / (2) Номер цели i	1	2	3	4	5
1	17,5	15	9	5,5	12
2	16	16,5	10,5	5	10,5
3	12	15,5	14,5	11	5,5
4	4,5	8	14	17,5	13
5	13	9,5	8,5	12	17,5

KEY: (1) Number of active means, j; (2) Number of target i.

Elements of the matrix are the probable damages in each variant of the target assignment. It is necessary to distribute active means on targets so that the total damage would be minimal. The Hungarian method is based on the principle according to which the optimum character of the solution (or solutions) of the problem of target assignment is not disrupted with a decrease (or increase) of all the row elements of the table (or its column) expressing the probable damages by the same value C. Finally, the target-assignment problem is represented by a matrix the elements of which are numbers 0 or 1, whereupon in each row (or in each column) there is only one unit (or zero) symoblizing the selected value of the probable damage.

The process of solving the target-assignment problem by the Hungarian method is divided into six stages.

Stage I. The obtaining of zeros.

Among the elements of each column of the table the smallest is selected. This smallest element is subtracted from all the elements of this column. A matrix is composed the elements of which are the differences

$$C'_{ij} = C_{ij} - \min_{i} C_{ij},$$

where the subscript i is the row, and the subscript j is the column (for example, C_{14} means the element belonging to row 1 and column 4). Having made stage I, we obtain Table 9.6. The described method makes it possible to obtain at least one zero in each column.

Table 9.6.

(1) Номер активного средства, j / (2) Номер цели i	1	2	3	4	5
1	13	7	0,5	0,5	6,5
2	11,5	8,5	2	0	5
3	7,5	7,5	6	6	0
4	0	0	5,5	12,5	7,5
5	8,5	1,5	0	7	12

KEY: (1) Number of active means j; (2) Number of target i.

In an example (Table 9.5) it follows to subtract 4.5 from all elements of the first column, 8 from all elements of the second column, and so on.

Stage II. Search for the optimal solution.

With the help of zero values C it is necessary to produce a search for the solution for which the total damage would have a zero value. If this is possible, then the optimum solution is found. Otherwise one should pass on to Stage III.

The search for the solution begins from an examination of that line (or those lines) which contains the smallest number of zeros. Let us enclose by a small square one of the zeros of this line and then delete the zeros which are located in the same row or same column as that of the enclosed zero. Then among the remaining rows let us find that one (or those) which contains the least of all zeros, and we will repeat the same process until we will be able to enclose no more new zeros by small squares.

In Table 9.6 at first by a small square the element C_{24} is distinguished, then C_{35}, C_{53} and finally C_{41}, and C_{42} is cancelled out although it would be possible to enclose C_{42} and cancel out C_{41} (Table 9.7).

Table 9.7.

j i	1	2	3	4	5
1	13	7	0.5	0,5	6,5
2	11,5	8,5	2	$\boxed{0}$	5
3	7.5	7,5	6	8	$\boxed{0}$
4	$\boxed{0}$	✕	5,5	12,5	7,5
5	8.5	1,5	$\boxed{0}$	7	12

Table 9.7 does not give the solution with a zero value. Actually, if we complete the target assignment by the selection of element C_{12}, then the obtained solution will be 7 + 0 + 0 + 0 + + 0 = 7. Consequently, transition to the next stage is necessary.

Stage III. Searches for the minimum set of rows and columns which contain all zeros.

The actions of stage III are fulfilled in this sequence:

a) marked by a cross (×) are all those rows which do not contain one zero enclosed by a small square;

b) each column containing a cancelled zero at least in one of the marked rows is noted;

c) each line having a zero enclosed by a small square at least in one of the marked columns is noted;

d) actions b and c are repeated in turn until there remain no lines or columns which can still be marked.

This process makes it possible to obtain the minimum number of rows and columns which contain all the deleted or enclosed zeros. In the example in question one should mark line 1; there are no columns which could be distinguished (Table 9.8).

Table 9.8.

i \ j	1	2	3	4	5	
1	13	7	0,5	0,5	6,5	X
2	~~11,5~~	~~8,5~~	~~2~~	$\boxed{0}$	~~5~~	
3	~~7,5~~	~~9,5~~	~~6~~	~~0~~	$\boxed{0}$	
4	$\boxed{0}$	$\xcancel{0}$	~~5,5~~	~~12,5~~	~~7,5~~	
5	~~8,5~~	~~1,5~~	$\boxed{0}$	~~7~~	~~12,5~~	

Stage IV. Completion of Stage III.

Let us delete each unmarked row and each marked column, which will make it possible to isolate the minimum number of rows and columns containing all deleted or enclosed zeros of the matrix.

In Table 9.8 it is necessary to delete lines 2, 3, 4 and 5 and not cross out the columns.

Stage V. The transfer of some zeros.

Let us examine the part of the matrix consisting of undeleted elements, and let us select smallest number in it. Let us subtract this number from elements of the undeleted columns and let us add to the elements of the lines.

In Table 9.8 elements of the row 1, the smallest element of which is equal to 0.5 are not lined out. Let us subtract it from all elements of columns 1-5, and let us add to elements of rows 2-5, which is actually equivalent in this case to the subtraction of 0.5 from elements of the first row (Table 9.9).

Table 9.9.

j / i	1	2	3	4	5
1	12,5	6,5	0	0	6
2	11,5	8,5	2	0	5
3	7,5	7,5	6	6	0
4	0	0	5,5	12,5	7,5
5	8,5	1,5	0	7	12,5

Stage VI. Obtaining of the optimum solution or transition to a new cycle.

According to Table 9.9 it is necessary to find the optimum solution according to rules of the second stage. If the latter is found, then the processes of solving the problem is finished. Otherwise all the stages are repeated.

As a result of the use of rule of the second stage data of Table 9.10 are obtained, and it contains new deleted or enclosed zeros.

The fulfillment of stage III makes it possible to mark row 5 and, consequently, column 4, on the basis of which there appears a mark opposite row 2. Let us line through rows 3 and 4 and columns 3 and 4; this will consist of stage IV (Table 9.11)

Table 9.10.

i \ j	1	2	3	4	5
1	12,5	6,5	$\boxed{0}$	X	6
2	11,5	8,5	2	$\boxed{0}$	5
3	7,5	7,5	6	6	$\boxed{0}$
4	$\boxed{0}$	X	5,5	12,5	7,5
5	8,5	1,5	X	7,5	12

Table 9.11.

i \ j	1	2	3	4	5	
1	12,5	6,5	$\boxed{0}$	X	6	×
2	11,5	8,5	2	$\boxed{0}$	5	×
3	~~7,5~~	~~7,5~~	~~6~~	~~6~~	$\boxed{0}$	
4	$\boxed{0}$	X	~~5,5~~	~~12,5~~	~~7,5~~	
5	8,5	1,5	X	7	12	×
			×	×		

Table 9.12.

i \ j	1	2	3	4	5
1	11	5	$\boxed{0}$	X	4,5
2	10	7	2	$\boxed{0}$	3,5
3	7,5	7,5	7,5	7,5	$\boxed{0}$
4	$\boxed{0}$	X	7	14	7,5
5	7	$\boxed{0}$	X	7	10,5

The transition to stage V makes it possible to find the smallest number (1.5) of the free part of the matrix (Table 9.11) and subtract it from elements of column 1, 2 and 5 and then add elements of rows 3 and 4. As a result the following matrix is obtained (Table 9.12).

At this time the new use of operations of stage II gives the zero solution entirely;

$$M = C^2_{41} + C^2_{52} + C^2_{13} + C^2_{24} + C^2_{35} = 0,$$

which is the one sought. In the initial matrix (Table 9.6) this solution corresponds to the damage $M = C_{41} + C_{52} + C_{13} + C_{24} + C_{35} = 4{,}5 + 9{,}5 + 9 + 5 + 5{,}5 = 33.5$ (arbitrary units).

The obtained solution can be represented by the matrix of the target assignment in the form of ones and zeros. Instead of the zeros, enclosed by small squares, ones are placed, and instead of all the remaining matrix elements - zeros.

Thus, the solution which leads to the minimum total damage of 33.5 arbitrary units, proves to be the following:

1st active means is assigned to target 4. Damage of 4.5 arbitrary units.

2nd active means is assigned to target 5. Damage of 9.5 arbitrary units.

3rd active means is assigned to target 1. Damage 9 arbitrary units.

4th active means is assigned to target 2. Damage of 5 arbitrary units.

5th active means is assigned to target 3. Damage of 5.5 arbitrary units.

Table 9.13.

j i	1	2	3	4	5
1	0	0	1	0	0
2	0	0	0	1	0
3	0	0	0		1
4	1	0	0	0	0
5	0	1	0	0	0

A similar calculation makes it possible to determine the maximum total damage of $M_{max} = 83.5$ arbitrary units. Between the best and worst solutions there is a difference of 50 arbitrary units and there are an additional 118 intermediate solutions (not necessarily differing from each other essentially; thus, for instance, the solutions C_{12}, C_{21}, C_{35}, C_{44}, C_{53} and C_{12}, C_{21}, C_{35}, C_{42}, and C_{54} are equivalent to each other).

This algorithm of the solution of the target assignments problem based on Hungarian method is simply implemented in the form of a program of operation of the computer, since in this case it is necessary to carry out only two operations: subtraction and comparison.

CHAPTER 10

COMBAT EMPLOYMENT OF AUTOMATIC CONTROL SYSTEMS IN AIR DEFENSE FORCES

A number of the technical solutions accepted in contemporary automatic control systems [ACS] (ACY) abroad can be of interest, since it shows the means of the automation of control of combat operations and also the basic directions of technical thought in the solution of problems which confront forces of antiaircraft defense.

This chapter is devoted to questions of the practical realization of the ACS arising during the armament of air defense forces of armies of the capitalist states.

10.1. Basic Functions Being Fulfilled by the Combat Crews of Control Posts Equipped with ACS

The center of the control system of combat operations of forces of antiaircraft defense is the control post, on the combat crew of which is placed the responsibility for the timely detection of the air enemy, the recognition of aerial targets within limits of its combat area, the warning about an air raid, the bringing into combat readiness of all means of antiaircraft defense, and the control of these means during the repulsing of an air raid.

In accordance with this on the control posts the combat crews fulfill the following basic functions:

- gathering of information from the subordinate means of detection of the air enemy;

- recognition of the detected aircraft;

- processing of the entire incoming information in accordance with the program composed in advance;

- preparation of possible variants of the use of combat means for repulsing a raid of the air enemy;

- clear representation of the aerial situation and state of the subordinate combat means;

- development of instructions for controlling combat means and the transmission of these instructions to the direct executors.

- development of instructions for controlling the means of the detection and transmission of these instructions to the executors;

- carrying out of the mutual exchange of information with neighbors;

- obtaining of instructions and information from the higher control post, and also the transmission of the necessary information to the higher control post;

- obtaining and transmission of information about the flights of own aircraft.

Naturally, the effectiveness of the combat operation of crews of the control posts is determined by the degree of automation of the fulfillment of these functions. Contemporary air defense systems are almost completely automated and the majority of the enumerated functions is accomplished with the aid of special equipment, the operating speed and accuracy of operation of the operators. The principle control post of the ACS is the computer (one or several), which fulfills the entire totality of operations on the processing of incoming information to the control post, the clear representation of the situation, the preparation of data for making a decision on the use of combat means, the development of instructions of control and guidance.

Information about the aerial enemy comes to the control post through channels of the transmission of data from radar sets and radar altimeters. Information from radar is taken either automatically or semiautomatically. The information is coded and transmitted only automatically.

Information about the state of one's own combat means, and also information of a secondary importance can come to the control post of system in a nonautomatic way and be introduced in a computer as needed by the operator with the aid, for example, of punch cards.

Since all sources of information and intermediate elements function asynchronously, i.e., the information comes to the control post with a different speed and discreteness, then the computer, in using the external memory units (storages), processes this information in sequence and at a speed determined by the program of the machine operation and in accordance with the significance of the individual forms of information.

The information being issued by the computer of the control post, can be divided into the following four basic forms:

- **information of the control being transmitted to units and subunits of the antiaircraft guided missile [SAM] (ЗУР);**

- information of the control being transmitted to subordinate control posts;

- information of the control of radar means of detection of the air enemy;

- information of the control (guidance) being transmitted to the side of the fighter interceptors found in the air.

The entire information is renewed (in the opinion of foreign specialists) with the required frequency and is transmitted to the line of communication automatically in digital code.

At the command post auxiliary information not requiring high speed of transmission is developed. This information is transmitted along command lines of communication without the use of high-speed automatic equipment.

The computer of the control post develops also the entire necessary information for the graphic representation of the air situation and state of one's own combat means. The latter is automatically represented on a different kind of signal panel and plotting boards connected with computer.

10.2. Removal and Initial Processing of Information on the Radar Connected with the ACS

The information obtained directly from the radar due to the presence of a large quantity of false targets greatly overloads the communication lines and the computer of the control post. Therefore, on the radar joined with the ACS the filtration and screening of surplus information are carried out. Furthermore, the useful information is converted into a binary code. The equipment intended for the fulfillment of the indicated operations is installed directly on the radar and is a specialized computer.

As a rule, this computer consists of two parts: equipment for the removal and processing of information entering from the radar and equipment providing measurement of the flight altitude of the targets.

First part of the computer obtains information from the radar and identification equipment (Fig. 10.1). It separates the useful signals from the noise and determines the coordinates and national identity of the targets. Then the entire obtained information is converted into a binary code. This part of the computer operates as a system of the mass servicing with limited waiting. In a complex air situation when not all the information can enter into the communication line, the obsolete information is erased. In this case it is assumed that the lost information will be renovated and transmitted in the next review.

Stored in the output device of the computer is all the information about the target which is transmitted to the control post of the ACS. Transmitted to the control post are the target azimuth and range, the storage time of the information, the criterion of recognition, the synchronizing pulses and a number of other data.

The second unit of the computer - the equipment which provides measurement of the flight altitude of the targets - is intended for:

- obtaining from the control post of inquiring instructions about the need for height measurement;

- for the development of the instructions of target designation to the radar altimeters;

- for development and transmission to the operator of the radar altimeter of visual information and data for making a decision and for the transmission of this decision to the control post of the ACS.

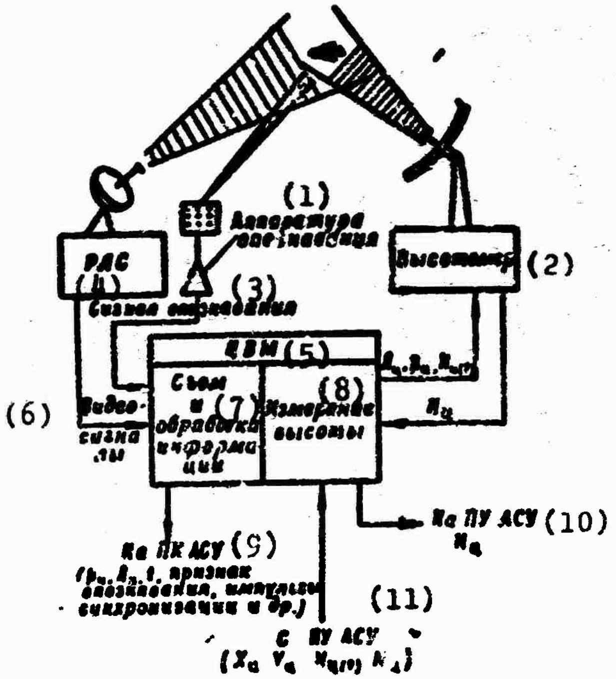


Fig. 10.1. Functional communication diagram of a radar unit with the computer for the removal and initial processing of information.
KEY: (1) Identification equipment; (2) Altimeter; (3) Identification; (4) Radar set; (5) Computer; (6) Video signals; (7) Removal and processing of information; (8) Height measurement; (9) To coordinate converter of ACS ($\beta_ц$, $Д_ц$, t criterion of recognition, synchronizing pulses and others); (10) To control post of ACS; (11) From control post of ACS.

The flight altitude of the target is determined thus. Coming from the control post of the ACS to the computer is an inquiry through the line of communication about the height of a specific target. Contained in the inquiry codes of the target position data $X_{ц}$, and $Y_{п}$, the supposed value of the height, and also the number of the radar unit and number of the target. Computer converts the rectilinear coordinates into polar $(\beta_{ц}, Д_{ц})$. The value of the azimuth of the target $\beta_{ц}$ is converted into the position of the shaft which is transmitted to the antenna of the radar altimeter, and values of target range $Д_{ц}$ and the supposed flight altitude $H_{п}$(?) are converted into voltages which are used in the radar altimeter for the development of the range and height gates. On the obtaining of inquiry instructions the altimeter turns the antenna around to the assigned azimuth and develops the range and height gates, which correspond to the indicated range and height. The fact that the antenna is directed toward the target it is possible to be convinced by the visual signals on the appropriate panel and by the glowing marks on the height indicator screen, which are intersected near the mark from the target the height of which must be measured.

The operator of the altimeter measures the height of the target by means of a control wheel, which moves the height mark on the screen. Located on the axle of the control wheel is a sensor of the scale altitude pulses. From this sensor the altitude value enters into the altitude register and then is automatically transmitted to the control post of the ACS.

10.3. Information Processing and the Control Combat Means of Control Posts of the ACS

It is advantageous to examine this question in the example of the operational center of the air defense sector equipped with the semiautomated "Sage" system. The operational center (Fig. 10.2) carries out the following missions:

- it processes information about the air situation (monitoring and filtration of input radar information, refinement and generalization of data of the air situation, final recognition of targets);

- it controls combat air defense weapons. (The control includes the distribution and selection of military weapons.)

Let us examine how these problems are fulfilled.

Information about the air situation enters into the operational center from early-warning radar and radar posts intended for the detection of low-altitude targets. This information passes the initial processing directly at the radar stations equipped with equipment of automatic removal, which does not exclude the possibility of the transmission of individual false targets and random noise and does not allow (especially in a complex air situation) the transmission of supplementary data on the target. In connection with this there appears the need for controlling and filtering out the input radar information in the operational center and produce secondary and tertiary information processing.

Radar information about targets which enters into the operational center in digital code consists of two numeric words following 10 μs after each other. The first word, ten-digit, contains the code of the radar number (4 digits) and codes of supplementary characteristics of the targets (5 digits). The second word, which is a twenty-two digit word, contains information about the azimuth of target (12 digits) and target ranges (10 digits).

The monitoring and filtration of the input information are conducted by operators of the monitoring of the input data with the aid of special equipment the logic circuits of which separate the entered message into parts and direct it along the appropriate channels for further processing. The subsequent processing consists in the deciphering of codes of the radar number and type of target in the conversion of codes of azimuth and range into the voltages corresponding to them.

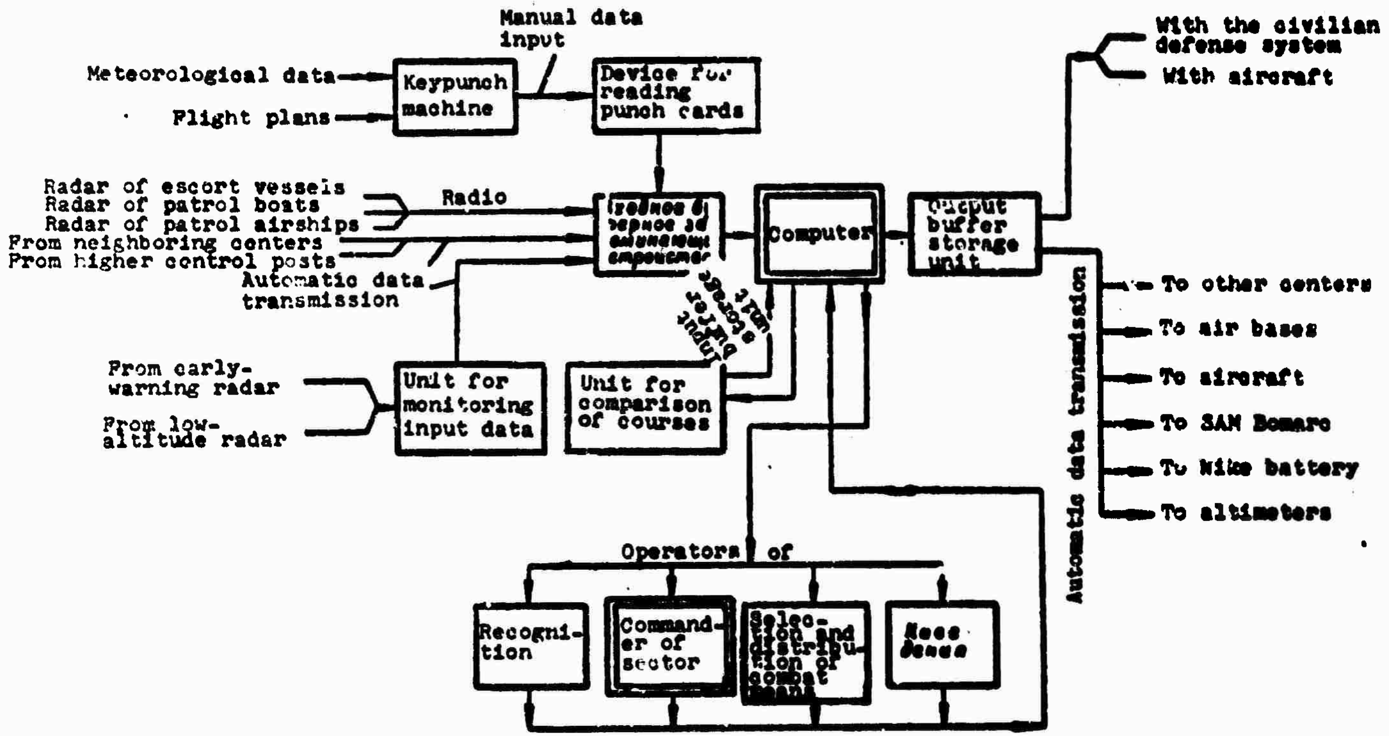


Fig. 10.2. Block diagram of the operation of the operational center of the air defense sector (USA).

Equipment for monitoring and filtration makes it possible to look over the entire input information about the air situation from any of the 15 radar sectors. Furthermore, it makes it possible to distinguish and issue for display 14 types of targets or objects. The presence of a selector of the type of target makes it possible for the operator to determine which types of targets are contained in the input information and filter the unnecessary signals.

Two operating modes of the equipment are provided:

- scanning on the indicator of one and any type of target;

- scanning of all targets simultaneously.

In the first mode the operator can automatically introduce the data on all examined targets into the computer, and in the second mode the selected targets are introduced by the operator semiautomatically. The selected information enters into the buffer storage of the computer.

The air situation is more precisely formulated and generalized by several operators. For each operator there is a panel of combat control with indicators using charactrons. From their post the operators manage all means of the detection of aerial targets and when necessary, for the refinement of the situation, are connected with the radar of their sector and with neighboring operational centers.

Further processing of radar information occurs in the computer, which converts the coordinates from the polar system into the rectangular system and ties them to the given operational center. After the converting of the coordinates from all the radars and the bringing of them into a single reference system, the newly entered data on the targets are compared with those earlier entered and extrapolated for the moment of the obtaining of new data

(Fig. 10.3). The computer, according to the nature of the trajectory, velocity and number of other criteria, determines the coincidence or noncoincidence of the data. If the data within the determined permissible limits coincided, then the computer, in plotting the course of the target, smooths it in accordance with the new data and extrapolates the course to the following period of the field of the radar scanning. This course enters into the storage (storage unit) of the extrapolated courses and indicators for display.

If the data does not coincide, a new mark is fixed as the possible new target, the data on it enter into the storage unit of the data on all unidentified targets, and the computer sends to the operator a signal in the form of specific mark on the screen about the noncoincidence of the characteristic or presence of a possible error. At the will of the operator further data processing is conducted automatically or manually.

With automatic processing the newly entered data on the target are compared with the earlier entered data on the unknown targets which are stored in the storage of the computer. If the data coincide, the computer introduces them into the storage as a target, confers a number on it and goes to automatic tracking. The target is illuminated on the indicators, and by a mark from the target primarily the altitude and all the remaining characteristics are measured. With the noncoincidence of the data the computer stores them in its storage during two-three scannings of the radar, comparing them with the foregoing scannings, and if noncoincidence continues to take place, the newly discovered mark is considered a new target and is transferred with the new number into automatic tracking. The previous data, which are stored in the storage, are erased from the storage of the computer.

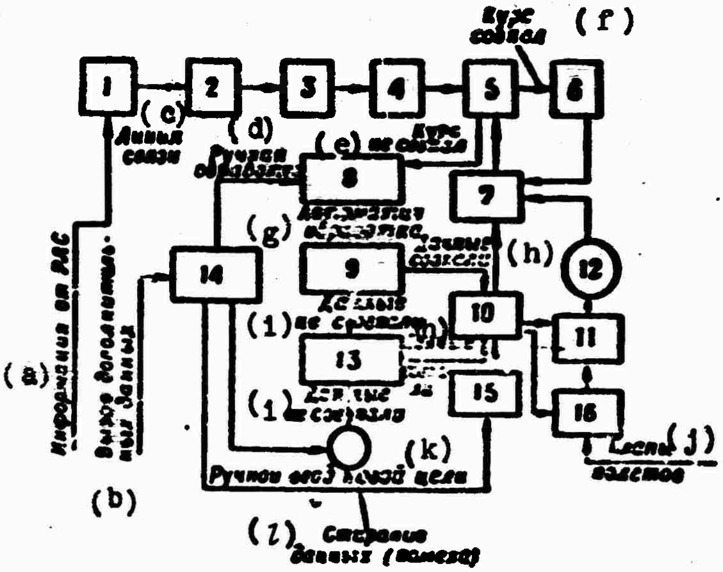


Fig. 10.3. Diagram of the processing of radar information in the operative center of the Sage system: 1 - Decoder of the operative center; 2 - Monitoring of input radar data; 3 - Input buffer storage; 4 - Converter of coordinates; 5 - Comparison of courses; 6 - Extrapolation of courses: 7 - Storage unit of the extrapolated courses; 8 - Storage unit of data on unidentified targets; 9 - Comparison with previous data of unidentified targets: 10 - Input of new target for automatic tracking; 11 - Comparison of data on the indicator; 12 - Identification operator; 13 - Comparison with data of unidentified targets in the future; 14 - Evaluation of the data on the indicator; 15 - Evaluation of data by the computer; 16 - Unit for comparing courses. KEY: (a) Information from radar; (b) Call of supplemental data; (c) Communication line; (d) Manual processing; (e) Course did not coincide; (f) Course coincided; (g) Automatic processing; (h) Data coincided; (i) Data did not coincide; (j) Flight Plans; (k) Manual input of new target; (*l*) Erasing of data (noise).

With the manual method of processing the entire radar information about targets is scanned on the indicators. For an analysis the operator can request from the computer or the radar supplementary information about the flight altitude, course and speed of target, induce information about the flight plans of his aircraft, and so on. The information requested by the operator is represented on the indicator. Operator's task is reduced to the comparison of the already available and newly obtained data and an analysis of the observed picture of the air situation. The operator makes the decision about the presence of a target or noise, erases the noise signal, introduces the target position data into the storage of the computer.

Besides the refinement of the air situation, the operators of the battle station generalize it. When the target is tracked by several radars, on the indicators several marks can appear from it. Generalization of the marks can be carried out by an operator manually. The computer is capable of analyzing and generalize the marks about the target automatically.

With a complex air situation the information can be enlarged both manually (by the operators) and automatically (the computer).

After the refinement, enlargement and generalization, the entire information is transmitted to the group of operators of the identification, who with the help of identification equipment on the radar identify the target. From their own (military) aircraft on the scopes different signs (for example, crosses, numbers or letters) are illuminated. A target can be identified for two reasons: either this is a foreign aircraft, or on this aircraft the responder of the identification equipment is absent.

With the appearance on the scope of a new unidentified target, on the control panel the buzzer sounds and a signal light ignites. On this instruction the operator puts the computer into the mode of the comparison of data from this target with the flight plans previously introduced into the computer with the help of a keypunch machine. The flight plan contains data on the course, speed, altitude of the flight and the supposed time of arrival of the aircraft into definite points. Data on the unidentified targets enter into the unit of the comparison of courses, where they are compared with the known flight courses of their own aircraft. With the agreement of the courses the computer qualifies the target as its own aircraft and marks it on the indicator by the appropriate sign, and with noncoincidence it issues to the indicator the sign of a "foe" (letter H in the log book). Furthermore, the information about this purpose transmits itself on the screens of the operators of the selection of military equipment.

The entire information about opponent's targets and also about the unidentified targets enters into the indicators of the post of the distribution of combat means and guidance. The senior operator, by choice and distribution of combat means, tracks the distribution of the targets between the individual operators and achieves the general monitoring. Operators of the selection of combat means is responsible for the rapid and most rational distribution of active means at targets. Operators of guidance directly guide the assigned active means at the target or control the guidance if it is achieved automatically.

The means between the operators are distributed by the computer, taking into account the place of the determination of the targets and their position relative to the active means of the sector. For the designation of the weapon to each target, the computer analyzes the state of all the active means of the sector. Primarily the readiness of the interceptors is checked. After this, depending on the characteristics of the target, the guidance method is selected. Then the computer determines the position of the interception point, the flight time of the interceptor to this point, and the quantity of fuel necessary for the carrying out of the combat mission. By comparing this quantity of fuel with the available reserve, the computer determines the correspondence of the range of the interceptors and the target range. By comparing results of the calculations, the computer selects the possible variants of the use of active means with minimum time for the interception and determines which of the operators of the selection of the combat means there will be transmitted this target and versions of solutions for it.

On the indicator of the operator of the selection of combat means the tactical situation and variants of the solutions processed by the computer are represented (Fig. 10.4). For all computations on the evaluation of the different active means and calculations of variants of solutions the computer spends 0.05 seconds. Having the indicator with the applied situation before himself, the operator makes a decision for the selection of those or different means of interception. In this case he can agree with one of the variants proposed by the computer or partially change it or propose his own variant. The operator introduces his decision into the computer by means of pressing the appropriate knob.

After the selection of the combat means, the control of them for the target kill of the enemy is placed on the guidance operators.

The computer transmits along the communication line to the airfield or to the control of the SAM a signal about the takeoff of interceptors or about the launching of a SAM. After detection by radar of the interceptor which took off, the latter is taken by the computer for automatic guidance, and the computer begins to solve the problem of guidance and also selects the guidance operator responsible for this operation of interception, sending to him the forewarning signal. On the scope selected by the computer of the guidance operator in the form of conventional symbols, the entire information necessary for guidance is represented. During interception the computer continuously calculates the necessary data for control of the interceptor, taking into account the maneuver of the target, weather conditions and a number of other factors.

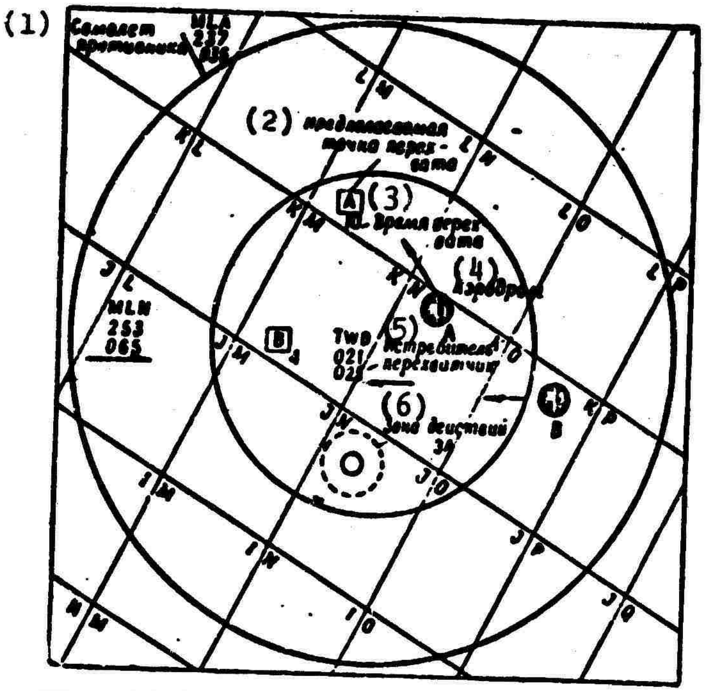


Fig. 10.4. Scope of the operator of the selection of combat means with a display of variants of decisions of the computer.
KEY: (1) Aircraft of the enemy; (2) Proposed interception point; (3) Time of interception; (4) Airfield; (5) Fighter interceptor; (6) Zone of coverage.

Instructions of guidance can approach the edge of the interceptor by several methods:

- automatically, directly to the autopilot;

- automatically in the form of the vocal signals formed in the computer which the pilot takes by radiotelephone;

- semiautomatically processing of instructions of guidance is automated, instructions of guidance are represented on a special signal panel in front of the operator of induction, and the latter reports them to the pilot on an ordinary radiotelephone).

Guidance continues until the target appears in the visibility range of the airborne radar of the interceptor and is detected by it. On the termination of the operation on interception, the operator of guidance turns a switch on the control panel to the position "Return to base." With this computer it produces the optimum data for the return of interceptor to the landing airfield; the data transmit themselves to the edge of interceptor automatically or by the operator.

If for the interception of the target the antiaircraft guided missiles Nike is selected, then the computer of the operative center sends a target indication signal to the control post ACS of the SAM, according to which the computer of this very system solves the problem of target assignment and missile guidance to the indicated target.

10.4. Brief Survey of Some ACS of Capitalist Countries

The Fire Control System of SAM Batteries of Missile Master

The system Missile Master (Fig. 10.5) is intended for the fire control of the SAM's Nike Ajax, Nike-Hercules and Hawk. It should provide the antiaircarft defense of separate objects. The system Missile Master can operate both in conjunction with the "Sage" system and independently of it. The system can coordinate fire for up to 90 launchers and is intended for the execution of the following functions:

- obtaining information about the air situation from the Sage system or from its (local) radar sets;

- providing the commander and officers of the control of the necessary data for making decisions;

- coordinating the fire of batteries of SAM's while maintaining their right to select independently targets for bombardment.

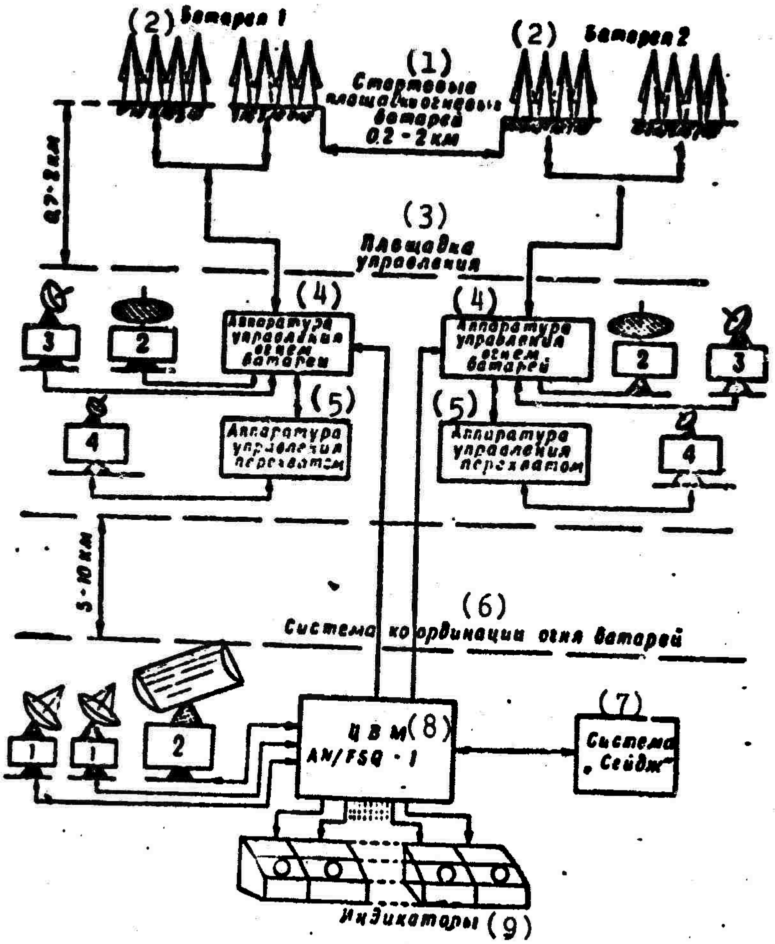

(2) Батарея 1
(2) Батарея 2
(1) Стартовые площадки огневых батарей 0,2 - 2 км
0,7-2 км
(3) Площадка управления
(4) Аппаратура управления огнем батареи
(4) Аппаратура управления огнем батарей
3
2
4
(5) Аппаратура управления перехватом
(5) Аппаратура управления перехватом
5 - 10 км
(6) Система координации огня батарей
1
1
2
ЦВМ (8) AN/FSQ - 1
(7) Система „Сейдж"
Индикаторы (9)

Fig. 10.5. Simplified diagram of the control of the complex of Nike SAM by means of the system Missile Master: 1 - Height finders; 2 - Detection radar; 3 - Target-tracking radar; 4 - Missile-tracking radar. KEY: (1) Launch area of fire batteries 0.2-2 km; (2) Battery; (3) Control area; (4) Fire control equipment of batteries; (5) Equipment of the control of interception; (6) System of the fire coordination of batteries; (7) Sage system; (8) Computer; (9) Indicators.

The system consists of the following basic elements:

- radar equipment for obtaining data on the aerial situation;

- control post with instrumentation for the automatic processing of data and their display;

- means of fire control of SAM batteries;

- system of the automatic transmission of instructions from the control post to the batteries and the reception of reports from the batteries to the control post.

The entire information about the aerial situation enters into the computer of the type AN/FSQ-1 installed at the control post. By means of the computer all operations on information processing, target assignment, output of information for display on indicators of the operators are automated. The conducting of fire at targets at the batteries is also automated, and responsibilities of the operators are reduced to monitoring the operation of the equipment and correction of the decisions of the computer.

At the control post of the system, besides the leader, who carries out general supervision, there are three groups of main operators:

- operators for the observation and tracking of targets;

- operators of the fire control of batteries;

- operators of identification.

Operators of the observation and tracking of targets observe the aerial situation on remote indicators and carry out identification of all targets in this area, the information about which is obtained from different sources of information, including the Sage system.

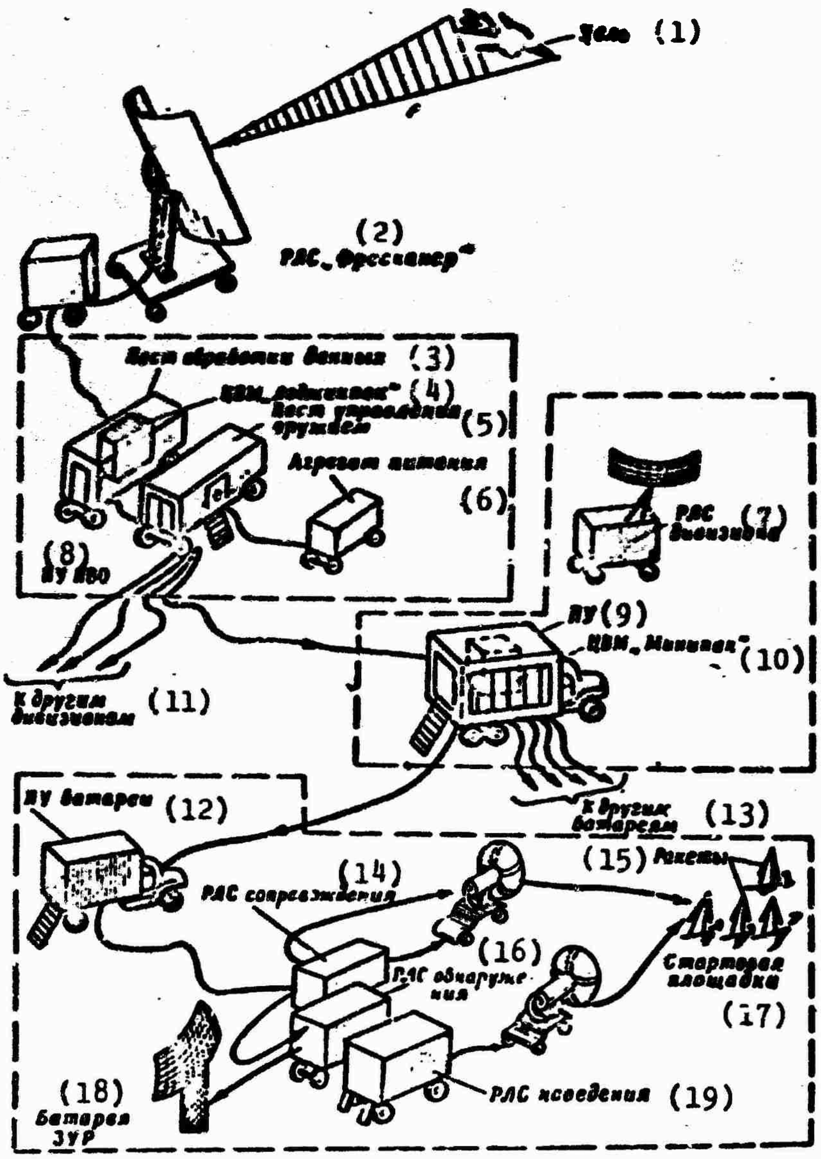


Fig. 10.6. Simplified circuit of the control of the military air defense system Missile Monitor.
KEY: (1) Target; (2) Radar Frescanner; (3) Post of data processing; (4) Computer Logicpac; (5) Control station of weapon; (6) Minipac computer; (11) To other batallions; (12) Control post of battery; (13) To other batteries; (14) Radar tracking station; (15) Missiles; (16) Detection radar; (17) Launch area; (18) SAM battery; (19) Guidance radar.

Operators of the fire control of batteries on sign indicators estimate the situation and control the selection of targets between the batteries, eliminating the excessive doubling of the fire.

Operators of identification, having obtained data on the national identity of the targets, control the movement of their own identified aircraft. The operators have the capability at any moment of stopping the conducting of fire at the aircraft erroneously accepted as being foreign.

Of the other air defense systems of the USA, one ought to mention the systems of Buic , Birdie, of Missile Monitor, military mobile systems Helilift and Mauler, and tactical systems "412L" and TACS.

From the English air defense systems most important are the Bloodhound and Fire Brigade systems.

In France the air defense system Strida-2 functions, and in Sweden - Stril-60.

The Buic system is the reduced and simplified version of the Sage system. The Buic system is assumed to be the doubling system of the Sage when the latter becomes inoperative.

The Birdie system is a considerably simplified version of the system Missile Master. The system is mobile and can be comparatively rapidly deployed.

The military air defense system Missile Monitor is intended for the antiaircraft defense of the combat area of field armies of the USA. The system (Fig. 10.6) is semiautomatic, mobile, and adapted for transportation on ground and in the air. In composition and operating principle it is in many respects similar to the air defense system Missile Master. The basic source of information in the system is the three-coordinate ($Д_ц$, $β_ц$ and $ε_ц$) radar "Frescanner." The removal of coordinates in the radar is semiautomatic by means of an electronic marker connected wtih a push-button mechanism on the axles of which are located the "shaft-number" convertors, which makes it possible to introduce target position data directly into the computer.

At the post of data processing there is installed a special-purpose computer "Logicpac," which processes all the information about the aerial situation and sends it to the display units of operators of the post of data processing and weapon control post. The computer estimates and distributes the targets by subdivisions entering into the system. When any battery obtains a target for fire action, drawn on the indicators of operators of the weapon control post is a glowing line, which connects the place of deployment of the battery with the target blip.

The control post of the battalion coordinates the actions of several batteries and is the connecting line between the control post of the air defense of the field army and batteries of SAM. The control post of the battalion obtains information about the aerial situation from its own radar and from the battery radar. For data processing of the aerial situation and the exchange of information between batteries, the computer Minipac (AN/TSQ-36) is used.

At the control posts of the batteries finally selected are the targets for bombardment, and a prefiring check and preparation of the SAM's, the target guidance and other operations connected with the delivery of fire are carried out.

The military system Helilift is intended for the fire coordination of SAM batteries placed in a large territory. This system is analogous to the Missile Monitor system.

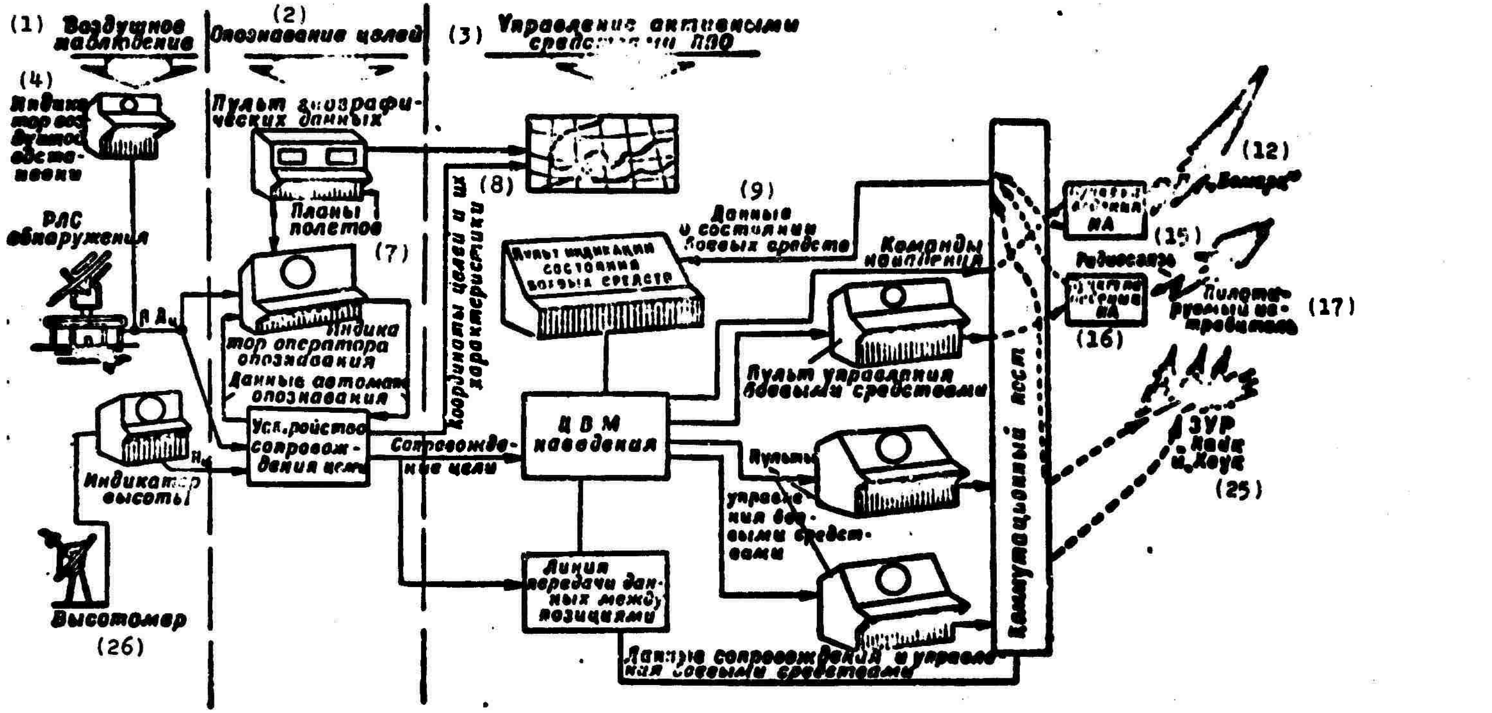

(1) Воздушное наблюдение
(2) Опознавание целей
(3) Управление активными средствами ПВО
(4) Индикатор воздушной обстановки
Пульт географических данных
Планы полетов
(7)
(8)
(9)
РЛС обнаружения
Индикатор оператора опознавания
Данные автоматического опознавания
Устройство сопровождения цели
Координаты целей и их характеристики
Сопровождение цели
ЦВМ наведения
Пульт индикации состояния боевых средств
Данные о состоянии боевых средств
Команды наведения
Пульт управления боевыми средствами
Пульты управления боевыми средствами
Линия передачи данных между позициями
Данные сопровождения и управления боевыми средствами
Индикатор высоты
Высотомер
(26)
Коммутационный пост
Пункт наведения ИА
Радиосвязь
(15)
(16)
Бомарк
(12)
Пилотируемый перехватчик
(17)
ЗУР Найк и Хок
(25)

Fig. 10.7. Simplified diagram of the control of the tactical air defense system "412L." KEY: (1) Aerial observation; (2) Target identification; (3) Control of active air defense weapons; (4) Indicator of aerial situation; (5) Console of geographical data; (6) Detection radar; (7) Flight plans; (8) Coordinates of targets and their characteristic; (9) Data on the state of combat weapons; (10) Instructions of guidance; (11) Guidance post for ИА; (12) Bomarc; (13) Indicator of the identification operator; Data of automatic identification; (14) Console of indication of the state of combat means; (15) Radio communication; (16) Guidance post for ИА; (17) Piloted interceptor; (18) Control panel of combat means; (19) Altitude indicator; (20) Unit for tracking targets; (21) Target tracking; (22) Guidance computer; (23) Consoles for controlling combat means; (24) Switching post; (25) SAM's Nike and Hawk; (26) Height indicator; (27) Line of data transmission between positions; (28) Data of the cracking and control of combat weapons.

The military mobile Mauler system is intended for the anti-aircraft defense of the advance units of ground forces and is a mobile compact system for controlling fire of a SAM. The entire system is mounted on one amphibious armored tracked personnel carrier.

The autonomous semiautomatic tactical system "412L" (Fig. 10.7) is intended for use by the mixed shock air groups of the Tactical Air Command of the Air Force outside the territory of the USA. The system provides the control of both the piloted and pilotless combat means. The system consists of the following basic elements:

- the radar for the assembly of the information about the aerial situation;
- equipment of the processing and display of data;
- communication equipment;
- equipment for the control of active air defense weapons;
- accessory equipment.

The radar sets detect and determine the coordinates and also identify the aerial targets. For the processing and display of the entire information there is used equipment which solves the problem of interception and issues the necessary data to the command for the selection of the most effective weapon of destruction. The combat readiness of all means is reflected on large signal panels with polychromatic indication. On a wall screen against the background of a map of the area, a general picture of the combat situation is represented. By means of consoles of combat control the selection of targets, guidance of fighter interceptors, their return to their bases after the attack, and the general coordination of combat operations are carried out.

The remaining air defense systems to a certain degree copy the described systems, differing only structurally and by the degree of the automation of control processes.

BIBLIOGRAPHY

1. **Анисимов Б. А., Четвериков В. Н.** Преобразование информации для ЭЦВМ. М., «Высшая школа», 1968.
2. **Барченков С. А.** Радиоэлектроника обучает, управляет и контролирует. М., Воениздат, 1966.
3. **Бубнов Г. П.** Автоматизация в ПВО. М., Воениздат, 1963.
4. **Венцель Е. С.** Введение в исследование операций. М., «Сов. радио», 1964.
5. **Дулевич В. Е., Коростылев Л. Е.** и др. Теоретические основы радиолокации. М., «Сов. радио», 1964.
6. **Каневский З. М., Финкельштейн М. И.** Флуктуационная помеха в обнаружение импульсных радиосигналов. М., Госэнергоиздат, 1963.
7. **Калашников Е. В.** Разведка и целеуказание. М., Воениздат, 1967.
8. **Каценбогев М. А.** Характеристики обнаружения. М., «Сов. радио», 1963.
9. **Клюев Н. Я.** Информационные основы передачи сообщений. М., «Сов. радио», 1966.
10. **Клюев Н. Ф.** Обнаружение импульсных сигналов посредством накопителей дискретного действия. М., «Сов. радио», 1963.
11. **Коряков В. Г.** Первичная обработка радиолокационной информации. Вторичная обработка радиолокационной информации. Третичная обработка радиолокационной информации. «Вестник ПВО», 1966, № 4, 6, 8.
12. **Крысенко Г. Д.** Современные системы ПВО. М., Воениздат, 1966.
13. **Кузьмин С. З.** Цифровая обработка радиолокационной информации. М., «Сов. радио», 1967.
14. **Куров В. С.** Боевое применение вычислительной информационной техники. М., Воениздат, 1964.
15. **Лоскутов В. И.** Вычислительная техника в процессах управления и контроля. М., «Знание», 1967.
16. **Лифшиц А. Л.** Кибернетика в военно-морском флоте. М., Воениздат, 1964.
17. **Романов А. Н.** Устройство автоматического съема данных. Цифровое измерение дальности. «Вестник ПВО», 1967, № 12; 1963, № 11; 1969, № 5.
18. **Самсоненко С. В.** Цифровые методы оптимальной обработки радиолокационных сигналов. М., Воениздат, 1968.
19. Справочник по основам радиолокационной техники. М., Воениздат, 1967.
20. Устройства ввода и вывода графической информации в ЭЦВМ (обзор зарубежного опыта). «Госстрой СССР», 1967.
21. **Фано Р.** Передача информации. Пер. с англ. М., «Мир», 1965.
22. **Финк Л. М.** Теория передачи дискретных сообщений. М., «Сов. радио», 1966.
23. **Ширман Я. Д., Голиков В. Н.** Основы теории обнаружения радиолокационных сигналов и измерения их параметров. М., «Сов. радио», 1963.
24. **Чайльд П. Ю.** Цифровая обработка данных РЛС. «Зарубежная радиоэлектроника», 1964, № 10.
25. **Чуев Ю. В.** и др. Основы исследования операций в военной технике. М., «Сов. радио», 1965.
26. **Хаббард Ю. В.** Цифровое автоматическое устройство для выделения радиолокационных данных. «Зарубежная радиоэлектроника», 1964, № 7.
27. Электронные системы отображения информации (под ред. Дж. Ховарда). Пер. с англ. М., Воениздат, 1966.
28. **Яблонский А. И.** Машина принимает решение. М., «Знание», 1964.